

Fyfe, Cameron D. (2015) *Structural and biophysical characterisation of Escherichia coli alpha-2-macroglobulin and its interaction with penicillin binding protein 1C*. PhD thesis.

<https://theses.gla.ac.uk/6971/>

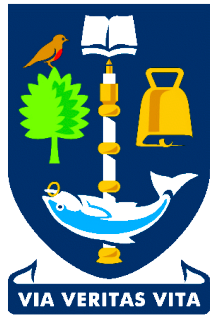
Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given



University of Glasgow

Structural and biophysical characterisation of *Escherichia coli* alpha-2-macroglobulin and its interaction with penicillin binding protein 1C

Submitted to the University of Glasgow for the degree of Doctor of
Philosophy

Cameron D. Fyfe, BSc (Hons), MRes

Submitted September 2015

Institute of Infection, Immunity and Inflammation
College of Medical, Veterinary and Life Sciences
University of Glasgow

Abstract

The alpha-2-macroglobulin superfamily consists of large multi-domain proteins that are activated by protease cleavage. One arm of this family consists of protease inhibitors that undergo a large conformational change upon protease cleavage, simultaneously physically trapping the cleaving protease and covalently linking to it via a thioester bond. However, there is little mechanistic understanding of how protease cleavage activates the conformational changes that lead to protease inactivation. These protease inhibitors are found in tetrameric, dimeric and monomer forms within eukaryotic blood/lymph fluid.

The recently described *Escherichia coli* alpha-2-macroglobulin (ECAM) is a periplasmic, inner membrane anchored protease inhibitor. The gene encoding ECAM, *yfhM*, is found within an operon alongside *pbpC*, which encodes penicillin binding protein 1C. These two proteins have been proposed to function in defence and repair against host proteases with ECAM acting to inhibit proteases that have breached the outer membrane and Pbp1C repairing damage to the peptide linkages within the peptidoglycan layer.

This thesis describes the structural and biophysical characterisation of ECAM and an investigation into the role of Pbp1C in ECAM function. In order to gain insight into the mechanism through which protease cleavage activates ECAM we used a combination of X-ray crystallography, small angle X-ray scattering and analytical ultracentrifugation to characterise the conformational changes that occur on protease cleavage. The X-ray structure of protease cleaved ECAM revealed a putative mechanism of activation and conformational change essential for protease inhibition. In this competitive mechanism, protease cleavage of the bait-region domain results in the untethering of an intrinsically disordered region of this domain which disrupts native inter-domain interactions that maintain ECAM in the inactivated form. Owing to the similarity in structure and domain architecture of ECAM and human α -2-macroglobulin, this protease-activation mechanism is likely to operate across the diverse members of this group. Further to this, it was shown that ECAM is processed in vivo, existing largely as truncated forms in growing *E. coli* cells. Interestingly, Pbp1C plays a key role in ECAM processing, with cell lacking *pbpC* showing an accumulation of full-length and dimeric forms of ECAM.

Acknowledgements

I would first like to thank my supervisor Dan Walker, you have provided the advice and support needed to develop as a scientist. The freedom to explore ideas and avenues has been an important aspect of supervision, although you knew when to reign me back in when I went off on a limb with little fruit!

My PhD would not have been the same in any other lab and I feel this is a lot to do with the four other guys I have been with in the Walker lab through most of it! Rhys, Kesha, Laura and Carla have been great for all science pub chats as well as stepping away from the science and having a great time! The addition of Sharon to the lab certainly made writing my thesis more fun and the future of the Walker lab looks like it's in good hands with Cat and Ploy! I'd like to thank Khedidja for starting me out on my protein purification ways, and for the quiche. Level 2 of the GBRC has become a great community of fun people that is only getting better with every addition. I would like to thank Olwyn Byron in particular for all your help with my biophysical work as well as always being up for having fun!

Without the financial support of the DTC in cell and proteomic studies I also wouldn't be where I am today. I'd like to thank Richard Burchmore for starting me off in Dan's direction and for the support you provided.

Finally I would like to thank my family for being there for me throughout my PhD, no matter the request for help, you've always been there!

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or at any other institution.

Cameron Fyfe

September 2015

Abbreviations

α 2M	alpha-2-macroglobulin
A2ML1	alpha-2-macroglobulin like protein 1
Amp	Ampicillin
ANA	Anaphylatoxin
anhMurNAc	1, 6-anhydro-N-acetylmuramic acid
AUC	Analytical ultracentrifugation
BA2M	Bacterial alpha-2-macroglobulin
BRD	Bait region domain
C3	Complement 3
C3a	Complement 3a
C3b	Complement 3b
C4	Complement 4
C4a	Complement 4a
C4b	Complement 4b
C5	Complement 5
C5a	Complement 5a
C5b	Complement 5b
CD	Circular dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CPase	Carboxypeptidase
CTMG	C-terminal macroglobulin domain
CUB	‘complement C1r/C1s, Uegf, Bmp1’ or ‘complement protein subcomponent domain’
Da	Dalton
dH ₂ O	Distilled water
DNA	Deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
ECAM	E. coli alpha-2-macroglobulin
EM	Electron microscopy

EPase	Endopeptidase
GlmS	Glucosamine-6-phosphate synthase
GlmM	Phosphoglucosamine mutase
HA2M	Human alpha-2-macroglobulin
IPTG	Isopropyl- β -D-thiogalactopyranoside
Kan	Kanamycin
kb	Kilobase pair
kDa	Kilodalton
L	Litres
LB	Lysogeny broth
LDAO	Lauryldimethylamine-oxide
Lpp	Lipoprotein
MBL	Mannose-binding lectin
mg	Milligrams
MG	Macroglobulin domain
ml	Millilitres
mRNA	Messenger RNA
M_w	Molecular mass
nm	Nanometres
OD ₆₀₀	Optical density at 600 nm
P	Poise
Pbp	Penicillin binding protein
Pbp1C	Penicillin binding protein 1C
PCR	Polymerase chain reaction
PZP	Pregnancy zone protein
RBD	Receptor binding domain
r.m.s.d	Root-mean-squared deviation
RNA	Ribonucleic acid
r.p.m	Revolutions per minute
R_g	Radius of gyration
SaA2M	<i>Salmonella enterica</i> ser. Typhimurium alpha-2-macroglobulin
SAD	Single-wavelength anomalous diffraction

SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SOMO	Solution modeller
SP	signal peptide
SV	Sedimentation velocity
TED	Thioester domain
TEP1	Thioester protein 1
TCEP	Tris(2-carboxyethyl)phosphine
TGase	Transglycosylase
TPase	Transpeptidase
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
UDP-MurNAc	Uridine diphosphate N-acetylmuramic acid
WT	Wild-type
µg	Micrograms

Table of Contents

Abstract	i
Acknowledgements	ii
Declaration	iii
Abbreviations	iv
Table of Contents	vii
List of tables	xi
List of Figures	xi
List of accompanying material	xiv
1 Introduction	1
1.1 Alpha macroglobulin superfamily	2
1.1.1 Eukaryotic macroglobulins	5
1.1.2 Alpha macroglobulin domain composition	8
1.1.3 Complement alpha macroglobulins	9
1.1.4 TEP1	13
1.1.5 Bacterial Alpha-2-macroglobulins	14
1.1.6 Thioesters in Gram positive bacteria	20
1.1.7 Thioester reaction	22
1.2 Penicillin binding proteins	23
1.2.1 Murein synthesis	23
1.2.2 Murein synthases	26
1.2.3 Murein hydrolases	30
1.2.4 Murein structure	31
1.3 Aims of the project	35

2	Materials and Methods.....	36
2.1	Reagents	37
2.2	Bacterial strains, plasmids and growth media	37
2.3	Preparation of chemically competent cells.....	40
2.4	Molecular biology.....	41
2.4.1	Polymerase chain reaction (PCR)	41
2.4.2	Oligonucleotide primers and sequencing	43
2.4.3	DNA restriction enzyme digest	44
2.4.4	Agarose gel electrophoresis.....	45
2.4.5	Agarose gel DNA extraction	45
2.4.6	DNA ligation	45
2.5	Transformation of chemically competent E. coli cells	45
2.6	Protein overexpression and purification	46
2.6.1	Cloning expression and purification of ECAM.....	46
2.6.2	Cloning expression and purification of Pbp1C.....	47
2.6.2.1	Solubilising Pbp1C	48
2.6.2.2	Refolding and purification of Pbp1C.....	48
2.6.3	Protein concentration	48
2.7	Protein concentration determination	48
2.8	SDS-PAGE.....	49
2.9	Western blotting	49
2.10	Mass spectrometry.....	50
2.11	Circular dichroism (CD) spectroscopy	50
2.12	Reaction of ECAM with proteases.....	50
2.13	Reaction of ECAM with methylamine	51
2.14	Analytical ultracentrifugation (AUC)	51
2.15	Protein crystallisation and data collection	52

2.16	Small angle X-ray scattering (SAXS).....	54
3	Overexpression, purification and characterisation of ECAM.....	56
3.1	Introduction.....	57
3.2	Results	58
3.2.1	Expression and purification of ECAM.....	58
3.2.2	ECAM is in a monomer-dimer equilibrium in solution	62
3.2.3	Small angle X-ray scattering of ECAM and comparison with high resolution models	68
3.3	Discussion	73
4	Crystallisation and structure model building of protease cleaved ECAM	76
4.1	Introduction.....	77
4.2	Results	77
4.2.1	Crystallisation of porcine elastase activated ECAM.....	77
4.2.2	Phasing and model building of porcine elastase activated ECAM.....	82
4.2.3	Overall structure of protease-activated ECAM.....	87
4.2.4	Interaction of the BRD with CTMG	91
4.2.5	BRD cleavage induced conformational shift	94
4.2.6	Thioester bond activation of α 2Ms.....	97
4.2.7	Biophysical characterisation of ECAM reacted with chymotrypsin	99
4.2.8	Small angle X-ray scattering of ECAM reacted with chymotrypsin and comparison with crystal structure of ECAM reacted with porcine elastase	102
4.3	Discussion	106
5	The role of Penicillin-binding protein 1C in the function of ECAM.....	112
5.1	Introduction.....	113
5.2	Results	115
5.2.1	In vivo analysis of ECAM	115

5.2.2	Role of other Pbps on ECAM processing.....	117
5.2.3	Pbp1C expression, refolding and purification.....	119
5.3	Discussion	123
6	Concluding remarks	126
7	References.....	131
	Publications	145

List of tables

Chapter 2

Table 2-1 <i>E. coli</i> strains and knockouts	37
Table 2-2 Protein expression vectors	39
Table 2-3 Complementation vectors	40
Table 2-4 Amplification PCR reaction	41
Table 2-5 Amplification PCR reaction components.....	41
Table 2-6 Mutagenesis PCR reaction	42
Table 2-7 Mutagenesis PCR reaction components.....	42
Table 2-8 <i>yfhM</i> (ECAM) primers	43
Table 2-9 <i>pbpC</i> (Pbp1C) primers	44
Table 2-10 ECAM partial specific volumes, buffer densities and viscosities calculated using SEDNTERP for the analysis of SV data	52

Chapter 3

Table 3-1 Experimental and computed sedimentation coefficient (S)	67
Table 3-2 Experimentally determined and computed ECAM biophysical properties	70

Chapter 4

Table 4-1 Data collection and refinement statistics (single-wavelength anomalous diffraction) for protease cleaved ECAM	84
Table 4-2 Movement of domains upon elastase activation	95
Table 4-3 Experimental and computed sedimentation coefficient (s) for ECAM reacted with proteases .	101
Table 4-4 Experimentally determined and computed ECAM biophysical properties	105

List of Figures

Chapter 1

Figure 1-1 Comparison of domain composition and structures of members of the α -macroglobulin superfamily.....	4
Figure 1-2 Alpha-2-macroglobulin structures.	6
Figure 1-3 Complement activation pathways and the involvement of alpha macroglobulin proteins.	10
Figure 1-4 Crystal structure of C3, C3a anaphylatoxin, and C3b opsonin.	12

Figure 1-5 Proposed defence and repair mechanism of bacterial α -2-macroglobulin and Pbp1C.	15
Figure 1-6 Structural models of ECAM from SAXS and EM.....	17
Figure 1-7 Domain organisation and crystal structure of unactivated Salmonella α 2M.	19
Figure 1-8 The structure of <i>Streptococcus pyogenes</i> and its location in the pilus adhesin	21
Figure 1-9 Thioester reaction with surface amines.....	22
Figure 1-10 Murein sacculus synthesis.	25
Figure 1-11 Penicillin binding proteins and their domain structure	27
Figure 1-12 Schematic of the primary structure of peptidoglycan pentapeptide subunit.	32
Figure 1-13 Electron micrograph of murein sacculus of E. coli and the three dimensional structure of peptidoglycan.	34

Chapter 3

Figure 3-1 Purification of ECAM by Ni ²⁺ --affinity chromatography and size exclusion chromatography (SEC) with SDS PAGE.	59
Figure 3-2 Purification of ECAM in the presence of TCEP by Ni ²⁺ -affinity chromatography and size exclusion chromatography (SEC) with SDS PAGE.....	61
Figure 3-3 Raw absorbance of ECAM during sedimentation velocity experiment.	62
Figure 3-4 AUC analysis of ECAM showing a monomer-dimer mixed population.	63
Figure 3-5 AUC analysis of ECAM purified with TCEP showing a monomer-dimer mixed population.	64
Figure 3-6 Increase in dimer with concentration	65
Figure 3-7 Crystal structure of <i>Salmonella enterica</i> alpha-2-macroglobulin	66
Figure 3-8 Comparison of raw SAXS data for ECAM at high and low concentrations.	68
Figure 3-9 Raw SAXS data for ECAM with Guinier fit.	69
Figure 3-10 Distance distribution function p(r) analysis and Kratky plot for ECAM indicate a globular particle in solution.	70
Figure 3-11 Comparisons between ECAM experimental scattering and CRYSOLOG generated unreacted SaA2M monomer and dimer high resolution model curves.....	72
Figure 3-12 ECAM dimer has a larger conformation than SaA2M high resolution model.....	73
Figure 3-13 ECAM model of interaction with inner membrane within periplasm.	75

Chapter 4

Figure 4-1 Size exclusion chromatography of ECAM and ECAM reacted with porcine elastase.	78
Figure 4-2 Crystals from ECAM reacted with porcine elastase and their diffraction.	79
Figure 4-3 Elastase treated ECAM crystals on SDS-PAGE and LC MSMS peptide coverage from bands.	80
Figure 4-4 Stereo image structure within 1 σ electron density envelope.....	85
Figure 4-5 Cleaved thioester of protease-cleaved ECAM.	86
Figure 4-6 Crystal structure of porcine elastase cleaved Escherichia coli α 2M (ECAM).	89
Figure 4-7 Comparison between trimmed ECAM, SaA2M, and HA2M.	90
Figure 4-8 Bait region interactions with the CTMG domain.....	92

Figure 4-9 Sequence alignments of conserved motifs involved in thioester pocket and conformational activation.	93
Figure 4-10 Conformational shift of TED relative to MG6 domain of protease-activated ECAM.	96
Figure 4-11 Movement of residues within thioester upon protease activation of bacterial α 2M.	98
Figure 4-12 Purification by SEC of ECAM reacted with chymotrypsin.	99
Figure 4-13 AUC analysis of ECAM reacted with chymotrypsin.	101
Figure 4-14 SAXS data for ECAM reacted with chymotrypsin.	103
Figure 4-15 CRYSOLOG comparison between elastase activated ECAM crystal structure and chymotrypsin activated ECAM SAXS curve.	104
Figure 4-16 Crystal structure of elastase treated ECAM overlaid with SAXS envelope of chymotrypsin treated ECAM.	105
Figure 4-17 Comparison of the thioester protecting pocket in eukaryotic and bacterial α 2Ms.	107
Figure 4-18 Alignment of protease-activated ECAM structures.	109
Figure 4-19 Putative mechanism of protease entrapment and inhibition by ECAM.	110
<u>Chapter 5</u>	
Figure 5-1 Bacterial α -2-macroglobulin operons and a schematic representation of their expressed proteins localized to the bacterial envelope.	113
Figure 5-2 Wild-type phenotype loses ECAM cleavage upon <i>pbpC</i> knockout and returns upon complementation.	115
Figure 5-3 Strains lacking genes involved in the synthesis of peptidoglycan maintain a wildtype ECAM processing phenotype.	117
Figure 5-4 Chemical inhibition of Pbp1C in WT shows similar phenotypic characteristics seen in Δ <i>pbpC</i> 118	
Figure 5-5 SEC of LDAO refolded Pbp1C.	120
Figure 5-6 Circular dichroism spectroscopy indicates that Pbp1C has native secondary structure.	121
Figure 5-7 Refolded Pbp1C shows no processing of ECAM <i>in vitro</i>	122
Figure 5-8 Possible composition of bands seen by Western blot.	124
Figure 5-9 Model of how Pbp1C and ECAM function in live cells.	125
<u>Chapter 6</u>	
Figure 6-1 Availability of bait region for cleaving proteases.	128
Figure 6-2 Model of ECAM reacting with Pbp1C and proteases.	130

List of accompanying material

Movie 1.mp4

1 Introduction

1.1 *Alpha macroglobulin superfamily*

The alpha macroglobulin superfamily is composed of globular proteins from a wide range of species with varied function. There are two main functional subtypes described to date with the first involved in inhibition of proteases through the formation of a complex between the alpha macroglobulin and protease. The second is involved in defence against pathogens, with the alpha macroglobulin covalently binding to the pathogen resulting in lysis or clearance. Alpha macroglobulins that have a protease inhibitory role include the eukaryotic proteins alpha-2-macroglobulin ($\alpha 2M$), pregnancy zone protein (PZP), CD109, CPAMD8, A2ML1 and the more recently characterised bacterial $\alpha 2M$ (BA2M) that is found in a wide range of Gram-negative bacteria¹⁻⁶. The second type of alpha macroglobulin play an important role in the eukaryotic innate immune system including complement proteins C3, C4 and C5 as well as TEP1 which has been described as having an important role in arthropod innate immunity⁷⁻¹⁰. The two branches of the alpha macroglobulin family have common evolutionary origins and although functionally distinct have similar mechanisms of activation, via protease cleavage^{11,12}. All alpha macroglobulins have a high molecular weight of ~180 kDa and some form homomultimers, with monomers consisting of a conserved order of domains albeit with some structural variation between species and subtype^{2,13}.

Comparisons between eukaryotic and BA2M have highlighted the strong similarity in the domains present as well as particular motifs that are required to maintain the function of these proteins (Figure 1-1)¹⁴. The structural differences seen between the proteins highlight the change in function between organism, $\alpha 2M$ having many long loops that are important for forming the tetramer structure and the extra MG domains seen in BA2M are important for anchoring the protein to the inner membrane^{13,14}. One feature that is conserved in most alpha macroglobulins is the thioester bond within the thioester domain (TED) which is important for covalently binding to the proteins corresponding target, be it a pathogen for complement $\alpha 2Ms$ and TEP1 or the cleaving protease for $\alpha 2M$ and BA2M^{10,13,14}. Within unactivated $\alpha 2Ms$ the thioester is protected from hydrolysis by solvent by a hydrophobic pocket between the TED and the C-terminal macroglobulin domain (CTMG)¹⁴. The residues involved in the hydrophobic pocket protecting this

thioester from solvent hydrolysis have been shown to differ between eukaryotic or archaea and BA2Ms¹⁴. In the TEP1 structure there are three tyrosines and a methionine in the CTMG orientated towards the thioester^{10,14,15}. These residues although conserved in all eukaryotic and archaea α 2Ms, however are not conserved within BA2M¹⁴. Instead in BA2Ms conserved residues that are involved in formation of the hydrophobic pocket are located in the TED itself¹⁴.

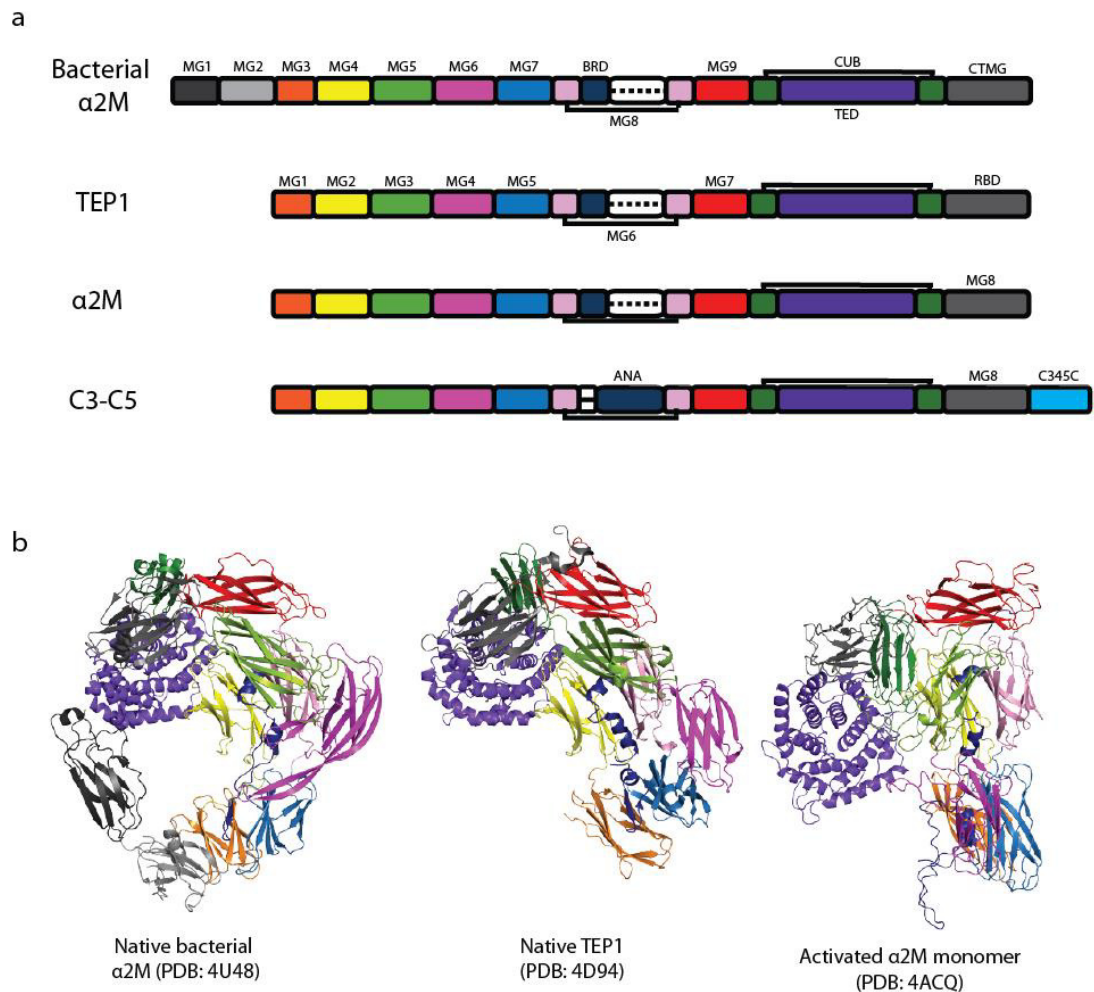


Figure 1-1 Comparison of domain composition and structures of members of the α -macroglobulin superfamily.

(a) Domain alignment of bacterial α -2-macroglobulin, thioester protein 1, α -2-macroglobulin and complement proteins C3, C4 and C5. (b) Crystal structures of native bacterial $\alpha 2M$, native TEP1 and methylamine activated α -2-macroglobulin aligned using macroglobulin domain 8 (bacterial $\alpha 2M$) and macroglobulin domain 6 (TEP1 and $\alpha 2M$). Note the similarity in orientation of domains between bacterial $\alpha 2M$ and TEP1 and difference seen in activated $\alpha 2M$. ANA, anaphylotoxin domain; BRD, bait region domain; CTMG, C-terminal macroglobulin domain; CUB, complement protein subcomponent domain; MG, macroglobulin domain; RBD, receptor binding domain; TED, thioester domain.

1.1.1 Eukaryotic macroglobulins

α 2M is a large soluble protein found in high concentration within human blood that functions as a broad spectrum protease inhibitor^{1,16,17}. This protein was first purified from blood in 1946 and was named alpha-2 due to the fraction of blood plasma in which it was found and macroglobulin due to its high molecular weight¹⁸. α 2M is produced mainly in the liver with blood plasma concentrations of as much as 2 mg ml⁻¹^{1,18,19}. This 720 kDa molecular weight protein complex is a tetramer made of subunits of 180 kDa, each comprised of 1451 amino acid residues^{16,17}. This tetramer has been described as a “dimer of dimers” with two monomers being connected by disulphide bonds and the secondary dimerisation interface forming via the interaction of bait regions (Figure 1-2 a)^{16,20}. Each of these tetramers have been found to inhibit proteases at a 1:1 or 1:2 ratio with variability found between protease types that are inhibited^{16,21}. The inhibition of proteases encompasses the four main classes of protease including serine, cysteine, aspartyl and metalloproteases via a “Venus flytrap” mechanism^{13,16,17}. This mechanism occurs where the α 2M is cleaved within its ‘bait region’ and this results in a large conformational rearrangement²². This rearrangement brings the normally buried β -cysteinyl- γ -glutamyl thioester into contact with the cleaving protease forming a covalent bond between the deaminated glutamine and surface lysines of the protease^{13,22}. Once the protease has been entrapped within α 2M it is no longer able to cleave large substrates due to steric hindrance of the active site. However, the protease active site remains intact and able to interact with smaller substrates and inhibitors^{1,16}. Without cleavage of the bait region, conformational activation of α 2M can be caused by chemical cleavage of the thioester bond with methylamine to form a γ -glutamyl methylamide. Treatment with small nucleophiles, such as methylamine, removes α 2M's ability to inhibit proteases^{17,23,24}.

Two forms of α 2M have been defined in reference to the speed at which they move during blue native electrophoresis with native α 2M moving slow and α 2M activated with protease or small nucleophile moving fast²⁰. Native and both chemically activated and protease-activated α 2M have been compared using electron microscopy (EM) and it has

been found that there is a distinct difference in shape between the native and activated forms, however, both types of activated $\alpha 2M$ are indistinguishable (Figure 1-2 b, c) ²⁵.

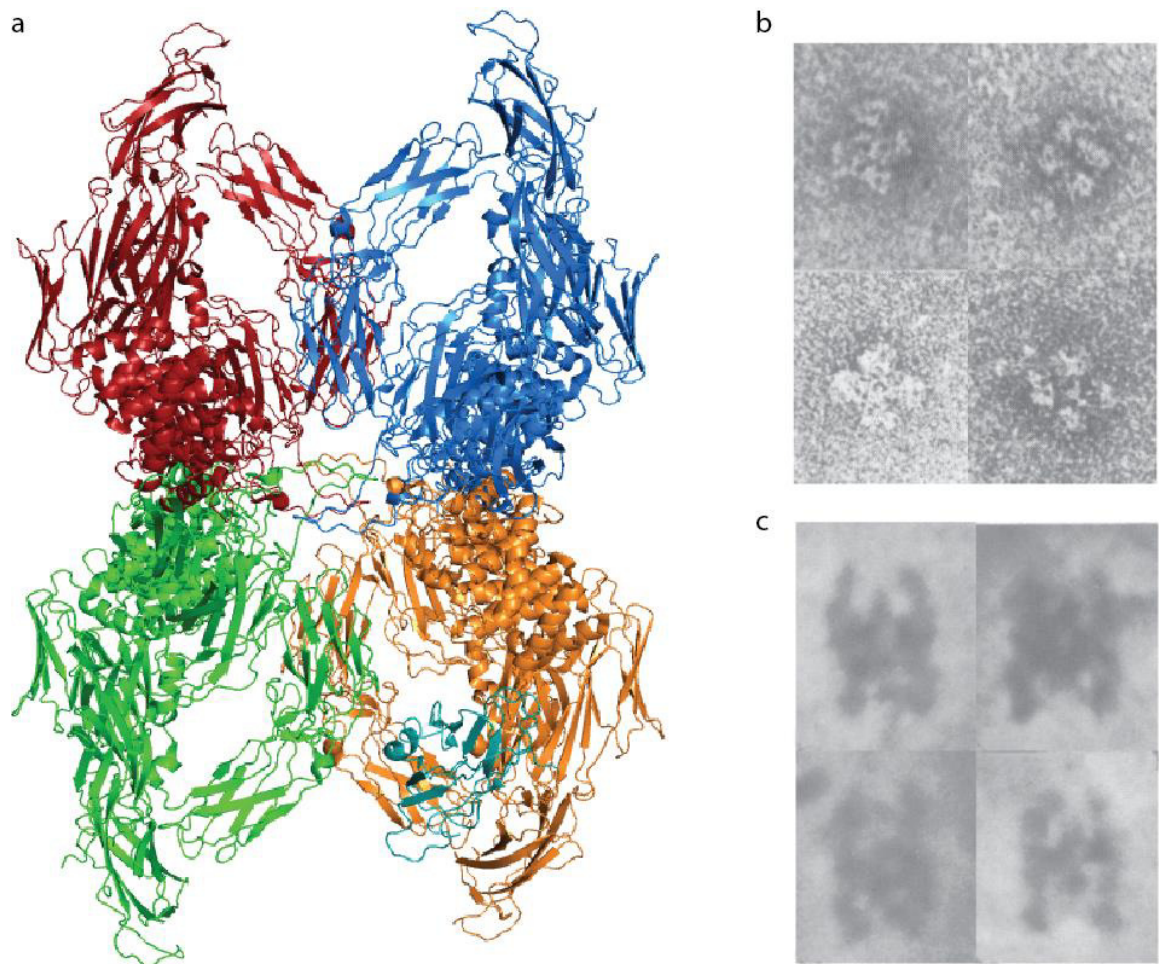


Figure 1-2 Alpha-2-macroglobulin structures.

(a) The crystal structure of tetrameric $\alpha 2M$ with subunits shown in red, blue, green and orange (PDB:4ACQ, Marrero et al 2012¹³). The domain shown in cyan is the C-terminal macroglobulin domain only present as part of the orange subunit. Electron microscope image of native (b) and trypsin activated (c) forms of $\alpha 2M$. Notice the similarity between the trypsin activated electron microscope image and the methylamine activated crystal structure. Electron microscopy images from Tapon-breaudiere et al²⁵.

Although there is no native crystal structure for $\alpha 2M$ published to date, the structure of methylamine activated $\alpha 2M$ has been determined to 4.3 Å resolution (Figure 1-2 a)¹³. Upon conformational activation the C-terminal macroglobulin domain (CTMG), also known as the receptor binding domain, becomes exposed at the surface of the tetramer. This CTMG domain has been shown to bind specific cell-surface receptors that result in endocytosis of the activated complex and degradation by lysosomes. This process results in clearance of activated $\alpha 2M$ from the blood within minutes of activation^{19,26}. $\alpha 2M$ has the ability to inhibit all four main classes of protease (serine, carboxyl, thiol, and metalloproteases) and has been shown to inhibit natively produced proteases such as chymase, a chymotrypsin-like protease produced by mast cells²⁷. $\alpha 2M$ can also inhibit proteases produced by bacteria and as such has been suggested in having a protective role against infection from pathogens²⁸. The ability of $\alpha 2M$ to bind various cytokines such as fibroblast growth factor, platelet-derived growth factor, nerve growth factor, interleukin-1 β , and interleukin-6 also indicates that it has an important role in the regulation of signalling²⁹⁻³³. With the ability to inhibit cytokines such as interleukin-6 it has further been suggested as having a role within chronic conditions such as Crohn's disease³³.

Various homologous protease and cytokine inhibiting alpha macroglobulins have been identified, some with specific targets. Pregnancy zone protein has been named as such due to the increased levels of expression seen during pregnancy². The levels of PZP increase from normal levels of 0.01-0.03 mg ml⁻¹ to 1 mg ml⁻¹ during the third trimester of pregnancy. With high sequence identity between $\alpha 2M$ and PZP of 72% it is thought PZP has a similar protease inhibitory role even though PZP exists as a dimer. PZP has been suggested as having a major role in controlling the activity of proteases that are released during cellular turnover that would be elevated during pregnancy. Alpha-2-macroglobulin like protein 1 (A2ML1) has strong similarity to $\alpha 2M$ and is expressed mainly in the epidermis⁵. A2ML1 lacks the cysteines involved in the disulphide bridges in $\alpha 2M$, involved in multimer formation, and is monomeric. Due to its expression in the epidermis A2ML1 is thought to play an important role in the shedding of the outer skin.

1.1.2 Alpha macroglobulin domain composition

The overall N-terminal to C-terminal domain architecture of α 2Ms consists of a series of seven macroglobulin (MG) beta sandwich domains that form a 'key ring' shape with a bait region spanning the centre^{13,14,34}. The macroglobulin domains are followed by a CUB ('complement C1r/C1s, Uegf, Bmp1' or 'complement protein subcomponent domain') arm that connects the TED to the beta 'key ring' with a final C-terminal MG 'receptor binding domain' (RBD)^{26,35,36}. The CUB domain that links the TED to the main 'key ring' is a sandwich of two beta sheets, half of which is before the TED and half is after leading to the final RBD¹³. The TED is mainly α -helical and contains a conserved cysteine-glutamine thioester present in most types of α 2M. This thioester forms a bond between the ϵ -cysteinyl- γ -glutamyl of the CXXQ motif^{23,37}. When the glutamic acid from the thioester comes into contact with lysine they form a covalent bond, linking the protease to the α 2M¹³. The final macroglobulin domain in human α 2M is involved in clearance of the protease cleaved activated form from the blood and as such is called the receptor binding domain²⁶. The bait region of α 2M is cleavable by a wide range of proteases suggesting a role in general housekeeping and regulation of proteases^{22,38,39}. However, some alpha macroglobulins do not carry a promiscuous bait region having a specified role in clearance of targeted proteases, cytokines or enzymes^{2,5}.

1.1.3 Complement alpha macroglobulins

The complement system involves a large network of proteins that results in the opsonisation or lysis of pathogens as well as mediating inflammation and eventual clearance of opsonised pathogens by phagocytosis⁴⁰⁻⁴³. The complement system obtained its name in the 1890s when it was found that a heat-labile protein found in the serum 'complemented' the killing of bacteria by heat-stable antibodies⁴². There are three complement proteins C3, C4 and C5 that are members of the alpha macroglobulin superfamily and each plays a major role in the complement system of human innate immunity. As part of the complement cascade these proteins are cleaved by various convertases and proteases (Figure 1-3). This results in a conformational change and activation of each protein forming C3b, C4b and C5b as well as releasing their respective C3a, C4a and C5a anaphylatoxins^{44,45}. This complement cascade can start through the binding of antibodies, mannose-binding lectins (MBL) or complement C3b to the surface of pathogens leading to three distinct pathways for complement cascade activation (Figure 1-3).

The alternative pathway is initiated when C3b opsonin binds to carbohydrates, lipids or proteins on the surface of pathogens. Upon C3b binding to a pathogen complement factor B binds to form C3bB that can be cleaved by complement factor D resulting in C3 convertase³⁴. C3 is always being hydrolysed to form C3b without requiring any other enzymes to cleave C3 to C3a anaphylatoxin and C3b. The constant generation of C3b allows the alternative pathway to begin without either of the other pathways being initiated.

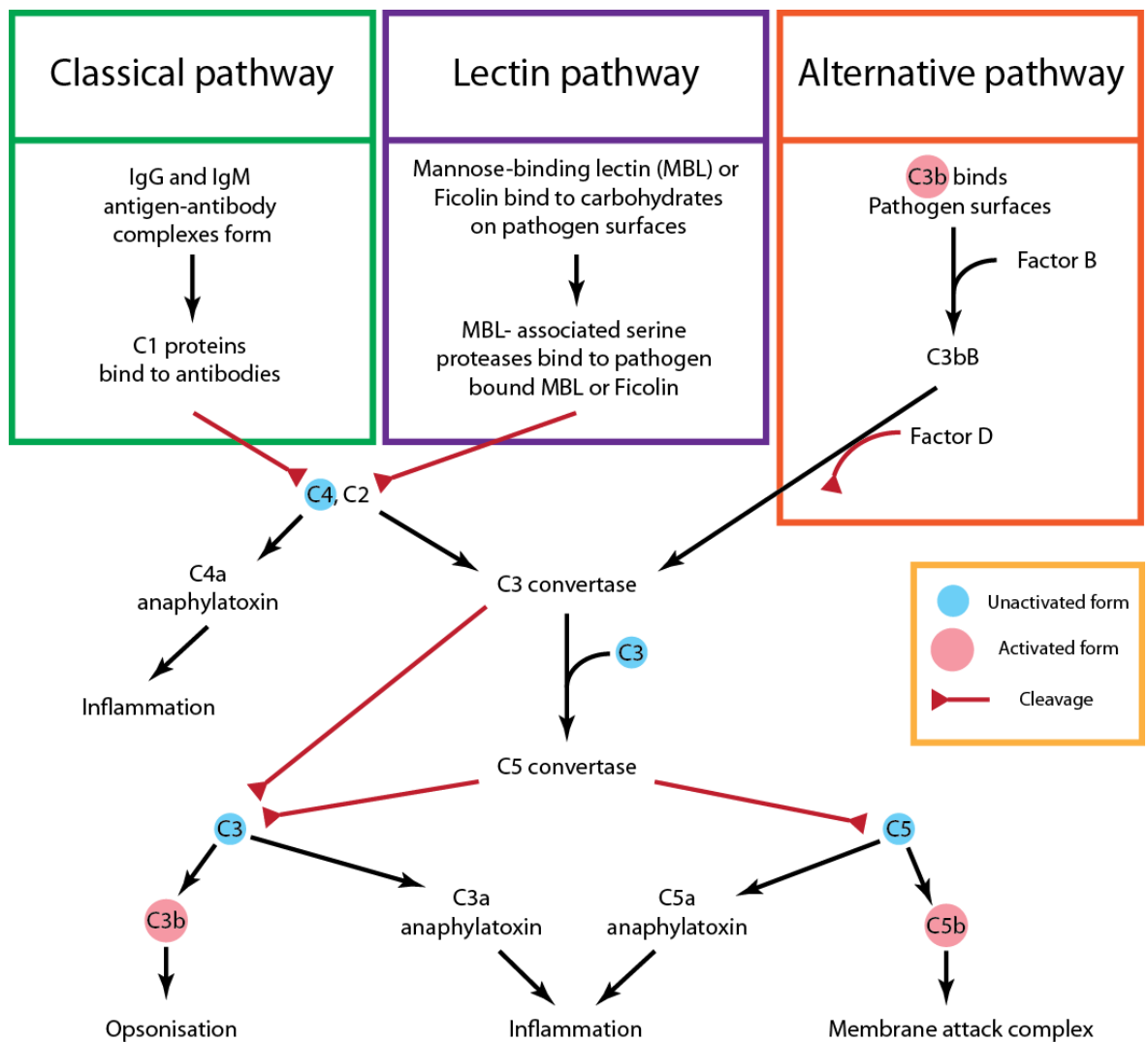


Figure 1-3 Complement activation pathways and the involvement of alpha macroglobulin proteins.

Complement C3 has multiple roles, it is important in activating the alternative pathway, in turning C3 convertase to C5 convertase and in the downstream opsonisation of pathogens. Complement protein C4 has a role in the classical and lectin pathway, upon cleavage it can bind C2a to become C3 convertase, which is able to cleave C3 and combine with C3 to become C5 convertase. Cleavage of C5 to C5b results in the generation of the membrane attack complex. The activation of C3, C4 and C5 results in the release of the anaphylatoxins C3a, C4a and C5a respectively leading to inflammation. Alpha macroglobulin proteins are highlighted in yellow and pink indicating inactive and activated forms respectively. The activated form of C4, C4b, becomes part of the C3 convertase complex resulting from the classical and lectin pathways. Adapted from Ricklin, Lambris; 2007 and Vidya Sarma, Ward;2012^{30,46}.

The mannose-binding lectin pathway is triggered by MBLs or ficolin binding to carbohydrates on the surface of pathogens⁴⁷⁻⁵². MBLs and ficolin are normally bound to MBL-associated serine proteases in circulation and upon binding to the surface of a pathogen these become activated and cleave C4 into C4a anaphylatoxin and C4b^{45,53}. The MBL-associated serine proteases are also able to cleave C2 and the resulting complex of C4b and C2a become the C3 convertase C4bC2a⁵⁴.

The classical pathway is activated with antibodies binding to antigens on the surface of pathogens. Complement C1q then binds the Fc (fragment , crystallisable) tail region of the bound immunoglobulin G or immunoglobulin M, with the subsequent binding of C1r and C1s proteins the new heteromultimeric protein becomes the C1 complex⁵⁵. The C1s subunit of the C1 complex is then able to cleave both C4, into C4a anaphylatoxin and C4b, and cleave C2 resulting in the C3 convertase C4bC2a.

All three pathways converge in generating C3 convertase which has the ability to cleave C3 into C3a anaphylatoxin and C3b opsonin (Figure 1-4). C3 convertase can also complex with C3b to become C5 convertase. C5 convertase is able to cleave both C3 into C3a anaphylatoxin and C3b opsonin as well as C5 into C5a anaphylatoxin and C5b⁵⁶. When considering the generation of C3b opsonin being one of the end points of the complement cascade and also the start of the alternative pathway there can be considered an amplification loop that leads to the generation C3b.

All three anaphylatoxins C3a, C4a and C5a have multiple similar effects on inflammatory response^{56,57}. A major role is their ability to activate chemotaxis in neutrophils and monocytes that bring these phagocytes closer to the complement activated cells to enhance clearance⁵⁸⁻⁶¹. These anaphylatoxins are also able to cause the release of histamine for activating neutrophils and causing vasodilation as well as well as causing cytokine release⁵⁸.

With the activation of C5 to C5b begins the generation of the membrane attack complex where complement proteins C6, C7, C8 form a complex with multiple copies of complement C9^{62,63}. The membrane attack complex subsequently lyses the attacked pathogen by inserting itself into the cells membrane and forming a pore.

Complement proteins C3, C4 and C5 share a similar domain composition although C5 does not contain a thioester⁸(Figure 1-1).

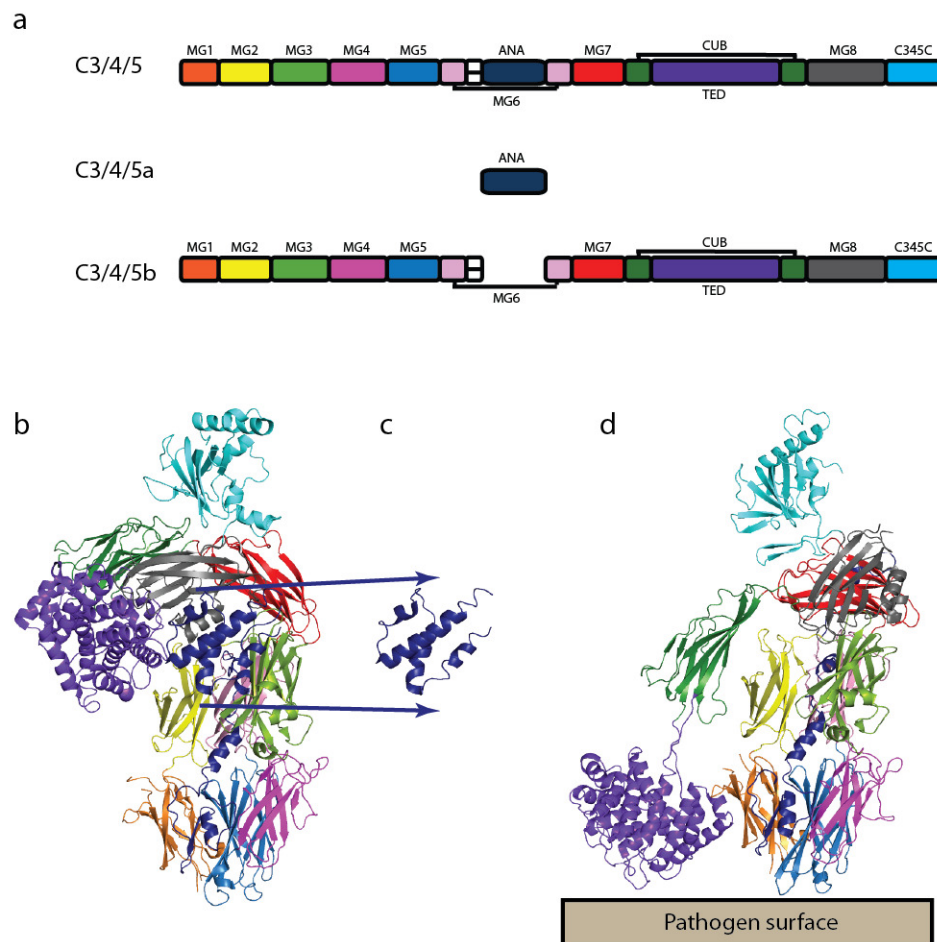


Figure 1-4 Crystal structure of C3, C3a anaphylatoxin, and C3b opsonin.

(a) A representation of the domains found in all complement alpha macroglobulins C3/4/5, C3/4/5a and C3/4/5b. (b) Crystal structure of C3 (PDB:2A73) before activation with domains coloured as shown in (a). (c) The anaphylatoxin C3a (PDB:2A73), with arrows indicating its location in C3, causes inflammation and chemotaxis to the site of complement activation. The movement of domains is shown in (d) upon cleavage of C3 to C3a and C3b (PDB:2I07)^{44,64}. Note the release of the TED so that it moves to orientate towards the pathogen surface for opsonisation. ANA, anaphylatoxin domain; CUB, complement protein subcomponent domain; MG, macroglobulin domain; TED, thioester domain.

Although there are striking structural similarities seen between $\alpha 2M$ and the complement proteins there are a number of features unique to complement proteins. All complement alpha macroglobulins have an anaphylatoxin domain in place of their bait region which is cleaved off of the remainder of the protein upon activation (Figure 1-4). Complement proteins also have a C345C domain which is important for binding of the proteases that are involved in cleaving off the anaphylatoxin domain. Another major difference between complement proteins and $\alpha 2M$ is the conformational change seen upon activation^{34,44,64}. For $\alpha 2M$, upon cleavage and activation, domains move to encapsulate the cleaving protease. In contrast in complement proteins, conformational reorientation of domains occurs bringing the thioester domain towards the surface of the pathogen to covalently bind to surface targets for opsonisation (Figure 1-4).

Links have been made between the evolutionary origins of $\alpha 2M$ and complement proteins^{11,65}. It is thought that $\alpha 2M$ predates the evolution of complement proteins and that the gradual evolution of the complement proteins predates the evolution of antibodies¹¹. There has been some suggestion that bacterial $\alpha 2M$ has been horizontally acquired from eukaryotic hosts as an adaptive mechanism of defence to foreign proteases⁶⁶.

1.1.4 TEP1

Thioester-containing protein 1 (TEP1) is a protein found in the malarial vector *Anopheles gambiae*, that similar to complement plays a role in innate immunity against pathogens^{10,15}. It has been shown to mediate the binding and killing of the malarial parasite *Plasmodium berghei* in conjunction with another heteroprotein, LRIM1-APL1C complex upon TEP1 cleavage and activation¹⁵. The LRIM1-APL1C complex is thought to stabilise the activated complex protecting the thioester to allow it to react with the pathogen. The domain architecture of TEP1 is very similar to other eukaryotic $\alpha 2M$ with close similarity also being noted between unactivated TEP1 and unactivated Saa2M¹⁰.

1.1.5 Bacterial Alpha-2-macroglobulins

α 2Ms have most recently been found in Gram-negative bacteria^{6,67}. Where eukaryotic α 2M are soluble and are usually found in multimer forms, BA2M's are membrane anchored lipoproteins with evidence suggesting they are monomeric^{14,67,68}.

Escherichia coli α 2M (ECAM) is the third biggest protein produced by *Escherichia coli* K-12 (BW25113) at 181 kDa, only surpassed only by a putative surface adhesin Yeej and ferredoxin reductase^{69–71}. ECAM is bound to the inner membrane located within the periplasm with an N-terminal anchor. The gene encoding ECAM, *yfhM* is found alongside the gene *pbpC* that encodes a penicillin binding protein. These two proteins have been suggested as having a linked function in defence and repair in which proteases that have breached the outer membrane are inhibited by ECAM, with PbpC repairing the damage caused to the peptide component of peptidoglycan⁶ (Figure 1-5).

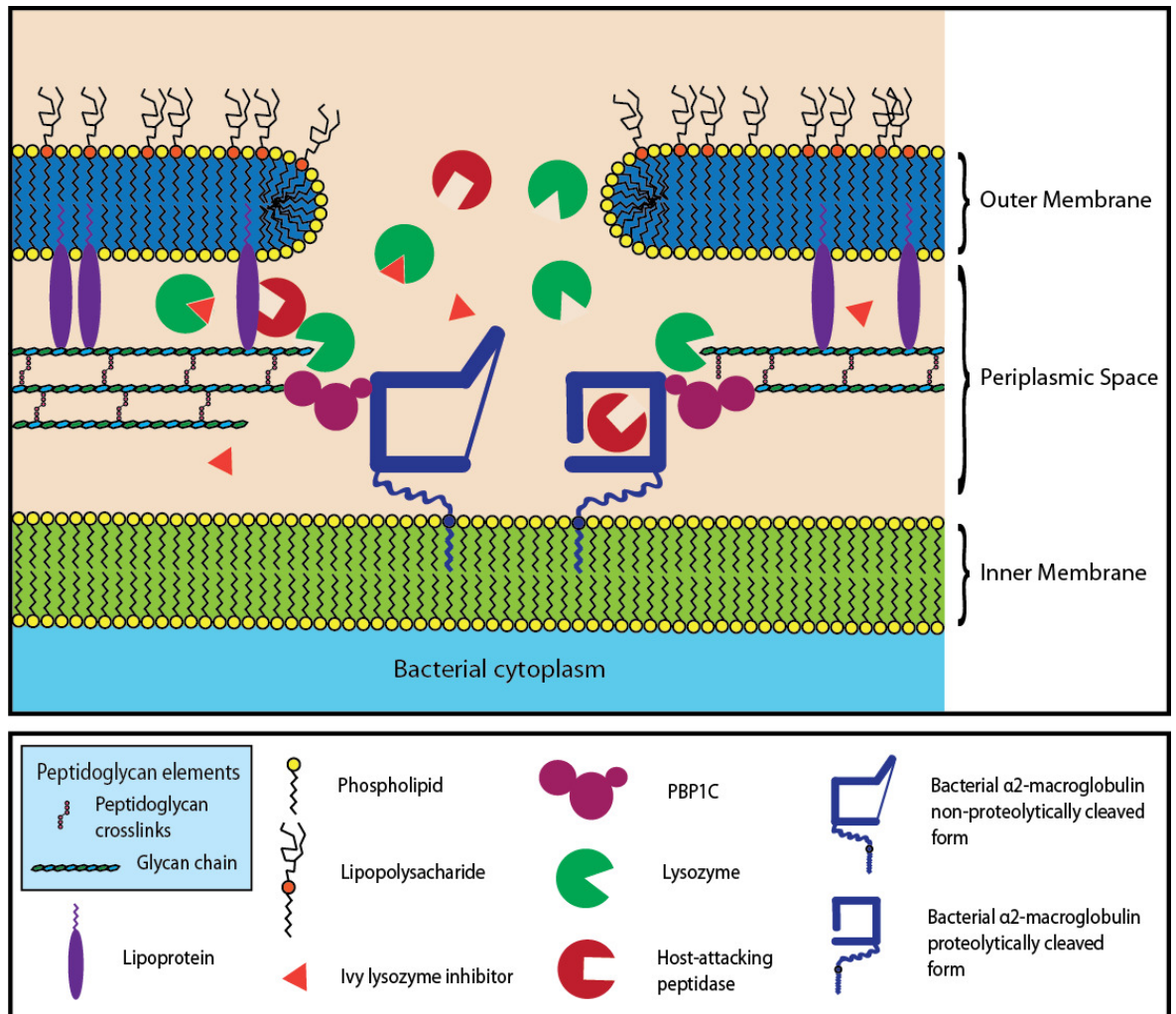


Figure 1-5 Proposed defence and repair mechanism of bacterial α -2-macroglobulin and Pbp1C.

Host antibodies, complement or antimicrobial peptides breach the outer membrane allowing host proteases and lysozyme entry to attack peptidoglycan elements. Host proteases are entrapped by bacterial α -2-macroglobulin and Ivy lysozyme inhibitors disable lysozyme allowing Pbp1C to repair the peptidoglycan. Figure elements not represented to scale and adapted from Budd *et al* 2004⁶.

Within the *E. coli* genome, as well as *yfhM*, there is a second ECAM encoding gene⁶. The second copy, *yfaS* is found in an operon with four other genes, *yfaA*, *yfaT*, *yfaQ* and *yfaP*. The genes within this operon, however, have had little investigation into their function. YfaS has only 15% amino-acid sequence identity with YfhM and does not contain a thioester and so would not be expected to covalently bind to proteases⁶. The *yfhM* and *yfaS* genes are conserved among a wide range of Gram-negative bacteria, although often with only one copy within a genome. A copy of at least one α 2M encoding gene has been found within alpha, beta, gamma, delta and epsilon proteobacteria with copies being found within various other phyla of bacteria⁶. An α 2M encoding gene has been found within the genome of archaea however these appear to have closer similarity to the eukaryotic form of α 2M rather than bacterial α 2Ms¹⁴.

The *Pseudomonas aeruginosa* genome has only one BA2M encoding gene, *magD*, which is within an operon that also contains the genes *magA*, *magB*, *magC*, *magE*, and *magF*⁶⁸. The proteins from this Mag operon have been suggested as having a role in bacterial pathogenicity due to their coregulation at the post-transcriptional level with other virulence factors⁷². *magA* mRNA has been shown to be a direct target of the RNA-binding protein RsmA that has been shown to bind and downregulate the expression of various genes important for synthesis of *P. aeruginosa* biofilms^{73,74}. Two notable operons that are downregulated by RsmA are *psl* and *pel* that are involved in exopolysaccharide^{72,74–76}. MagD has been suggested as not having direct interaction with the inner-membrane, but is bound via interaction with MagB⁶⁸. This complex is also thought to involve MagA and MagF, as they can be identified along with MagB via immunoprecipitation for MagD followed by proteomic analysis⁶⁸. The ability of MagD to interact with human neutrophil elastase suggests a similar function to YfhM as a protease inhibitor⁶⁸.

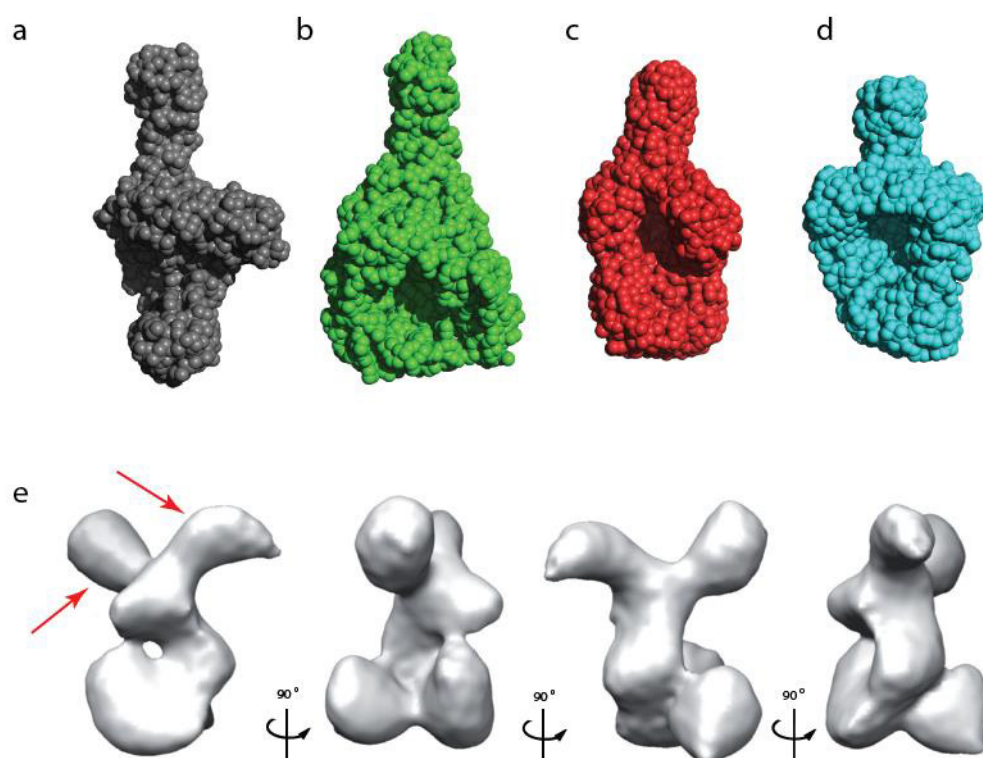


Figure 1-6 Structural models of ECAM from SAXS and EM.

Small angle X-ray scattering *Ab initio* models generated using GASBOR of (a) native untreated ECAM, (b) methylamine-treated (c) chymotrypsin-treated and (d) elastase-treated ECAM. (e) Isosurface representations of the 3D reconstruction obtained for the methylamine treated form of ECAM with red arrows highlighting regions of potentially greatest flexibility. Figure produced from figures shown in Neves *et al* 2012⁷⁷.

Structural characterisation of ECAM was first attempted by small angle X-ray scattering (SAXS) and electron microscopy (EM) where native ECAM was compared with ECAM activated with methylamine, or the proteases elastase and chymotrypsin^{68,77} (Figure 1-6. a, b, c, d respectively). These forms were compared due to expected similarity between bacterial α 2M and human α 2M in conformational activation via small molecule and proteases. Within this work it was found that the envelopes generated after reaction with methylamine and chymotrypsin and elastase showed distinct differences (Figure 1-6). Further work was performed using the *P. aeruginosa* MagD α 2M that appeared to show very similar envelope shape⁶⁸. Although both of these SAXS studies using ECAM and MagD found that there was a difference between native BA2M and methylamine reacted BA2M the recently published crystal structure of *Salmonella enterica* spp. Typhimurium α 2M (SaA2M) (Figure 1-7) showed that there was no conformational change on methylamine treatment¹⁴.

The high resolution model of SaA2M displays distinct similarity to α 2M¹⁴. Although SaA2M was described as being monomeric it was noted that a crystallographic dimer was present. All of the domains seen within human α 2M are present in SaA2M, however, there are an additional two N-terminal domains that anchor it to the inner membrane¹⁴. These two domains also form part of the dimer interaction seen within the crystal with contacts also found between these domains and the bait region. EM performed on methylamine activated ECAM alongside the earlier SAXS work showed a loop region and a flexible region, with similarities between the flexible region in this EM model and the C-terminus of complement⁷⁷ (Figure 1-6 e).

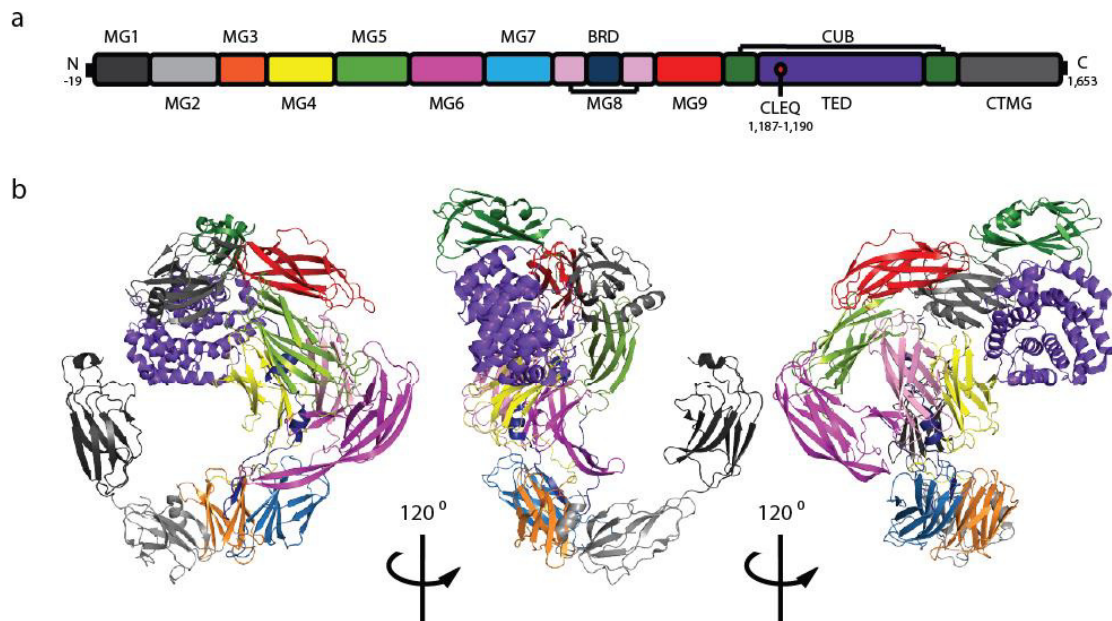


Figure 1-7 Domain organisation and crystal structure of unactivated *Salmonella* α2M.

(a) The domain composition including macroglobulin domains (MG) (highlighted C-terminal macroglobulin domain), bait region domain (BRD), CUB domain and the thioester domain (TED) that contains the CLEQ thioester. The crystal structure of *Salmonella* α2M shown (PDB:4U48 from Wong, Dessen¹⁴) in (b) has been coloured according to the domain arrangement seen in (a).

1.1.6 Thioesters in Gram positive bacteria

Not only have thioesters been identified in alpha macroglobulins but also in Gram-positive bacterial pilus adhesins (Figure 1-8 a)⁷⁸. These exist as long macromolecular filaments called pili that are important for bacterial virulence (Figure 1-8 b)⁷⁸. These have been shown to allow bacteria to colonise host organisms and have been suggested as being viable for development of vaccines⁷⁹⁻⁸¹. The thioester found in gram positive pilin adhesins has been suggested as functioning in a similar fashion as complement proteins in reacting with surface amines⁸².

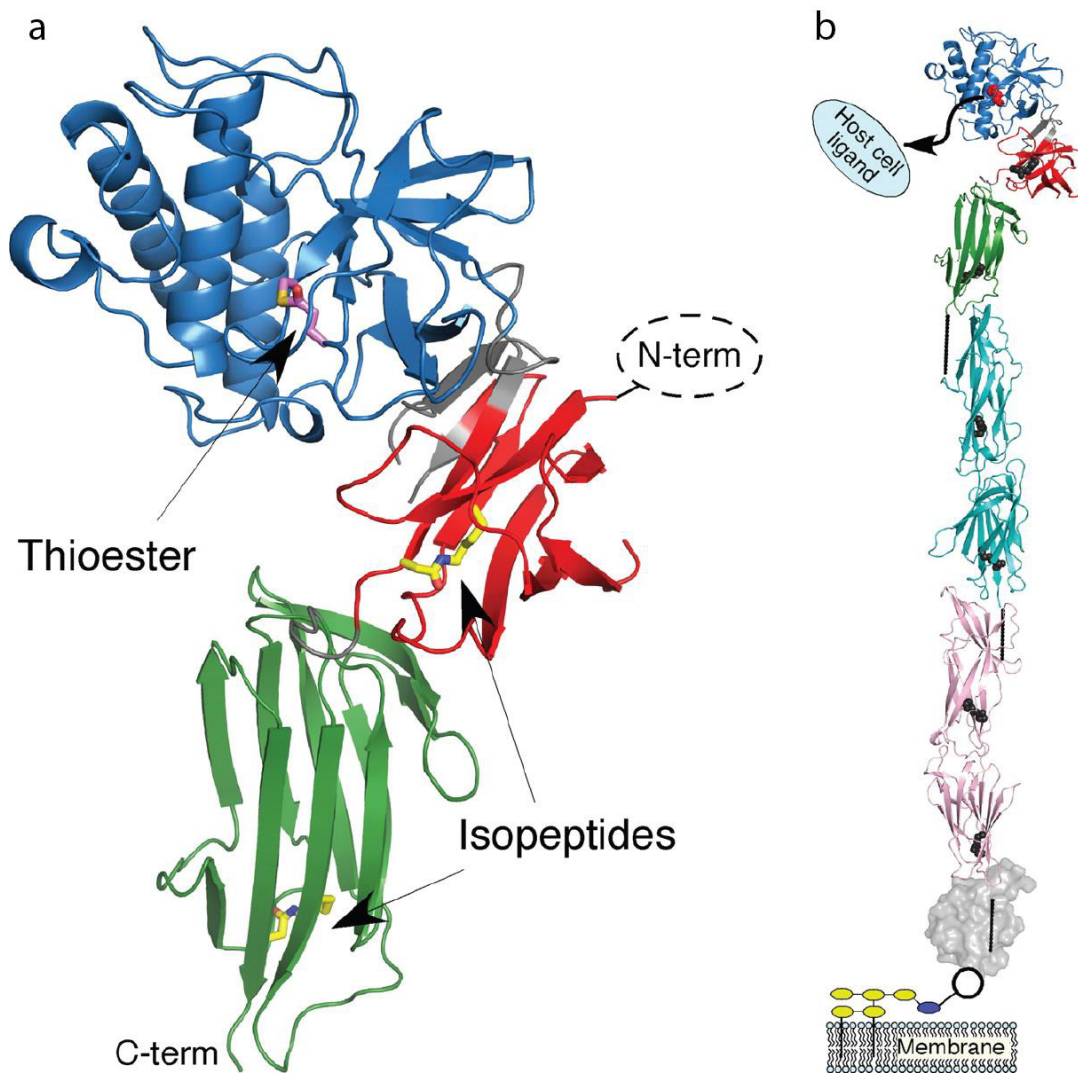


Figure 1-8 The structure of *Streptococcus pyogenes* and its location in the pilus adhesin

(a) Shown is a ribbon representation of the middle, top, and bottom domain coloured red, blue, and green, respectively; the termini are also labelled. Residues composing the thioester bond (Cys-426 and Gln-575) are shown with their carbon atoms in magenta. Residues composing the intramolecular isopeptide bonds (Lys-297 and Asp-595 in the middle domain, Lys-610 and Asn-715 in the bottom domain) are shown with their carbon atoms in yellow. Regions connecting the domains are shown in grey. The missing Spy0125-NTR (not expressed in the construct) would be positioned to the right of the middle domain. (b) A structural model of an intact pilus is shown. Spy0125 is coloured as above except the thioester and isopeptide residues are shown in red and grey. Two molecules of Spy0128 are shown (representing the polymerized pilus, cyan and pink with intramolecular isopeptide bonds coloured as above). Spy0130 is presented as a grey molecular surface. The location of the intermolecular isopeptide bonds are shown as solid lines. Yellow and blue ovals represent cell wall and cell wall precursors, respectively. Figure reproduced from Jonathan A. Pointon et al. J. Biol. Chem. 2010⁷⁸.

1.1.7 Thioester reaction

Alpha macroglobulin proteins require activation by proteolysis and domain movement, which exposes the thioester bond, whereas pilins do not require activation (Figure 1-9)^{14,37,82,83}. Alpha macroglobulins and pilins have a preference for amino groups however in C4b and C3, a His/Asp substitution near the thioester bond changes the substrate specificity and allows reaction with hydroxyl groups^{14,37,82,83}.

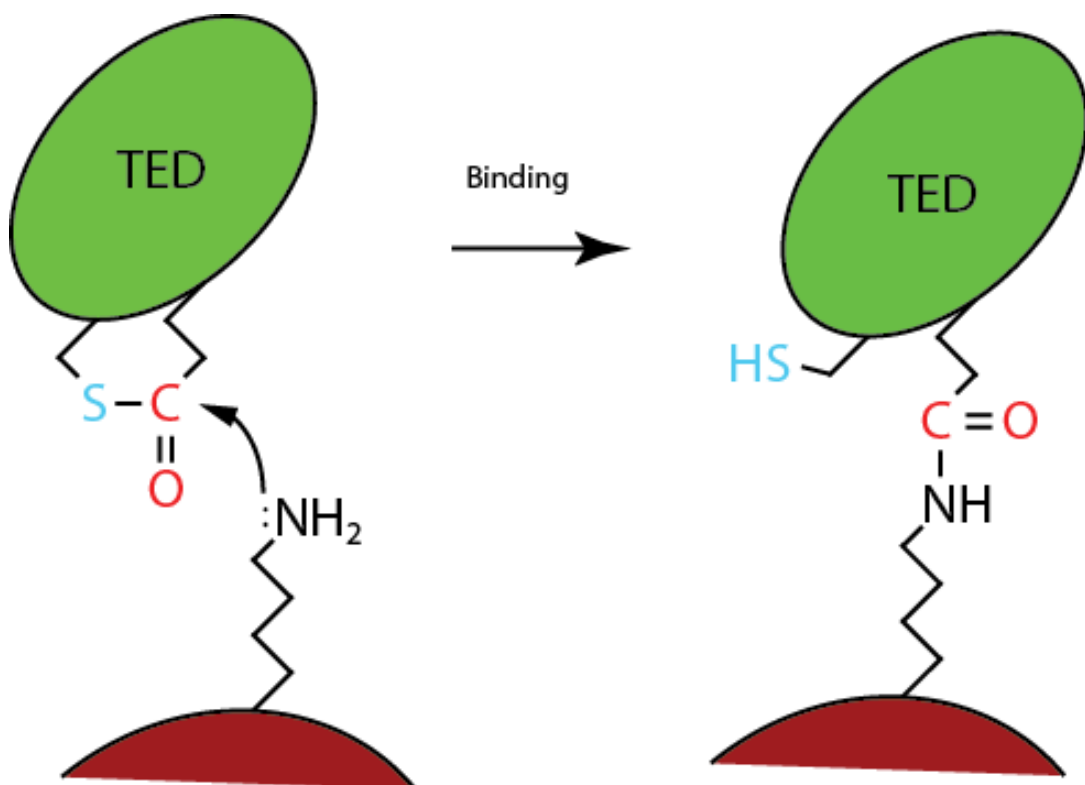


Figure 1-9 Thioester reaction with surface amines

Shown is a simplified scheme for the reactions of the thioester bonds in the thioester domain (TED, green) of alpha macroglobulin proteins (including complement, α 2M and BA2M) and pilin adhesins with surface amines (dark red). Figure modified from Linke-Winnebeck *et al* 2014⁸².

1.2 *Penicillin binding proteins*

Penicillin binding proteins (PBPs) are enzymes that catalyse the expansion of the murein sacculus by addition of a heterodimer glycan subunit with a pentapeptide side chain that is cleaved from lipid II^{84–88}. The name penicillin binding protein comes from their ability to covalently bind to penicillin as well as other β -lactam antibiotics⁸⁹. The interaction of β -lactams and PBPs is due to the similarity in structure between β -lactam antibiotics and the D-alanine-D-alanine termini of the pentapeptide sidechain found on peptidoglycan^{90–92}. During this interaction the β -lactam amide bond is cleaved forming a covalent bond with the serine residue in the PBP active site. PBPs have various functions in the synthesis and regulation of peptidoglycan structure. There are three main classes of PBP, Class A, B, and C. Class A PBPs are bifunctional transglycosylase/transpeptidases (TGase/TPase) murein synthases^{84,87}. Their TPase is the target of β -lactam antibiotics and functions in crosslinking pentapeptide sidechains between glycan chains^{84,87}. Class B PBPs are monofunctional TPases without the TGase activity found in Class A PBPs. Class C PBPs have several subclasses consisting of carboxypeptidases (CPases) and endopeptidases (EPases) that are involved in the regulation of peptidoglycan turnover^{84,87}.

1.2.1 *Murein synthesis*

Synthesis of murein subunits occurs within the cytoplasm of Gram-negative bacteria through the action of a number of enzymes. Biosynthesis begins with the conversion of uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) from fructose-6-phosphate by the sequential action of glucosamine-6-phosphate synthase (GlmS), phosphoglucosamine mutase (GlmM) and glucosamine-1-phosphate acetyltransferase and N-acetylglucosamine-1-phosphate uridylyltransferase (the final two processes performed by the GlmU bifunctional enzyme)^{92–97}. The next main step involves the conversion of UDP-GlcNAc to uridine diphosphate N-acetylmuramic acid (UDP-MurNAc) by the enzymes MurA and MurB^{98,99}. The addition of the pentapeptide sidechain is then catalysed sequentially by MurC, MurD, MurE and MurF^{100–103}. The anchoring of UDP-MurNAc-pentapeptide to an inner-membrane monophosphorylated undecaprenol is catalysed by

MraY to form lipid I¹⁰¹. The final step of addition of a GlcNAc to the lipid I to form a lipid II for transport to the periplasm is performed by MurG^{104,105}. Transport of lipid II from the cytoplasm across the inner membrane was long considered to be performed by either FtsW or RodA however recently it was discovered that a flippase enzyme MurJ is responsible for lipid II transport¹⁰⁶. Once transported across the inner membrane lipid II has undecaprenyl-diphosphate cleaved off and the new subunit is polymerised into the murein sacculus by murein synthases (Figure 1-10).

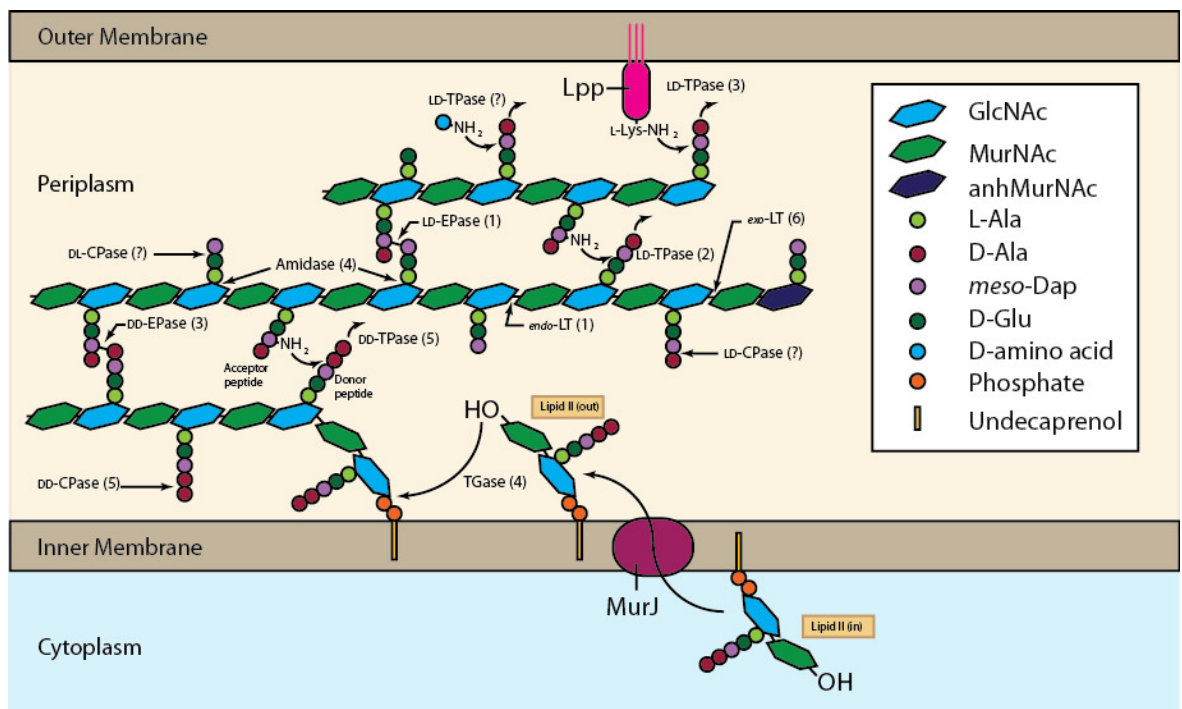


Figure 1-10 Murein sacculus synthesis.

Peptidoglycan subunits (Lipid II) are synthesised in the cytoplasm and transported to the periplasm by flippase MurJ and glycosyl chains are extended by transglycosylases (TGase). DD-transpeptidases (DD-TPase) further crosslink peptide sidechains to form the lattice of the peptidoglycan murein sacculus. Uncrosslinked peptides are cleaved by DD-, LD- and DL-carboxypeptidases (CPases), and crosslinks are cleaved by DD- and LD-endopeptidases (EPases). Peptides are removed from the glycan chain by amidases, and *exo- endo*-specific lytic transglycosylases (LTs) cleave the glycan chain to form 1, 6-anhydro-N-acetylmuramic acid (anhMurNAc) residues that are characteristic of the ends of glycan chains. The anchoring of the peptidoglycan sacculus to the outer-membrane is catalysed by LD-TPases forming crosslinks to outer-membrane lipoprotein (Lpp) as well as the catalysis of integration of unusual D-amino acids to the peptidoglycan. The number of known enzymes identified in *Escherichia coli* are shown in brackets although not all enzymes involved are known or characterised yet. GlcNAc, N-acetylglucosamine; meso-Dap, meso-diaminopimelic acid; MurNAc, N-acetylmuramic acid. Figure recreated and modified from Typas *et al* 2012¹⁰⁷.

1.2.2 Murein synthases

There are several types of enzyme that work together to synthesise the peptidoglycan network within the periplasm, including bifunctional TG/TPases, monofunctional TPases, and monofunctional TGases^{84,108}. There are two types of TPase enzyme within murein synthesis, DD-TPases and LD-TPases. DD-TPases form crosslinks between the meso-diaminopimelic acid of one sidechain and D-alanine of another^{109,110}. LD-TPases form crosslinks between the meso-diaminopimelic acids of two side chains¹¹¹.

Class A Penicillin binding proteins are large, ~80 kDa, inner membrane bound periplasmic proteins involved in the synthesis of the murein sacculus^{112–114}. These have both a TGase domain for crosslinking new subunits into the peptidoglycan and TPase domain for crosslinking peptides between chains. Three members have been identified within *Escherichia coli*, Pbp1A, Pbp1B and Pbp1C with a double knockout of Pbp1A and Pbp1B being lethal.

Penicillin binding protein 1A (Pbp1A) is encoded by the gene *ponA* (*mrcA*) whose single deletion results in minimum effect on cell growth or morphology¹¹⁵. This protein has been shown to be a major target for β -lactam antibiotics as strains lacking Pbp1B are more tolerant to β -lactams. An outer membrane activating protein LpoA, which interacts directly with Pbp1A, is required to activate Pbp1A function¹¹⁰. Pbp1A also functions alongside Class B PBPs to form a complex anchored to cytoskeletal elements within the cells cytoplasm through Mre and Rod proteins^{116,117}. These proteins have a role in dispersed elongation of the cell as well as presepation elongation near the central point where septation occurs¹¹⁸. Pbp1A has a small cytoplasmic region and spans the inner-membrane with a TGase domain at the inner-membrane interface showing the five conserved motifs seen in TGase domains (Figure 1-11). The C-terminal domain has DD-TPase function for crosslinking peptides within the peptidoglycan (Figure 1-10)^{112,119}.

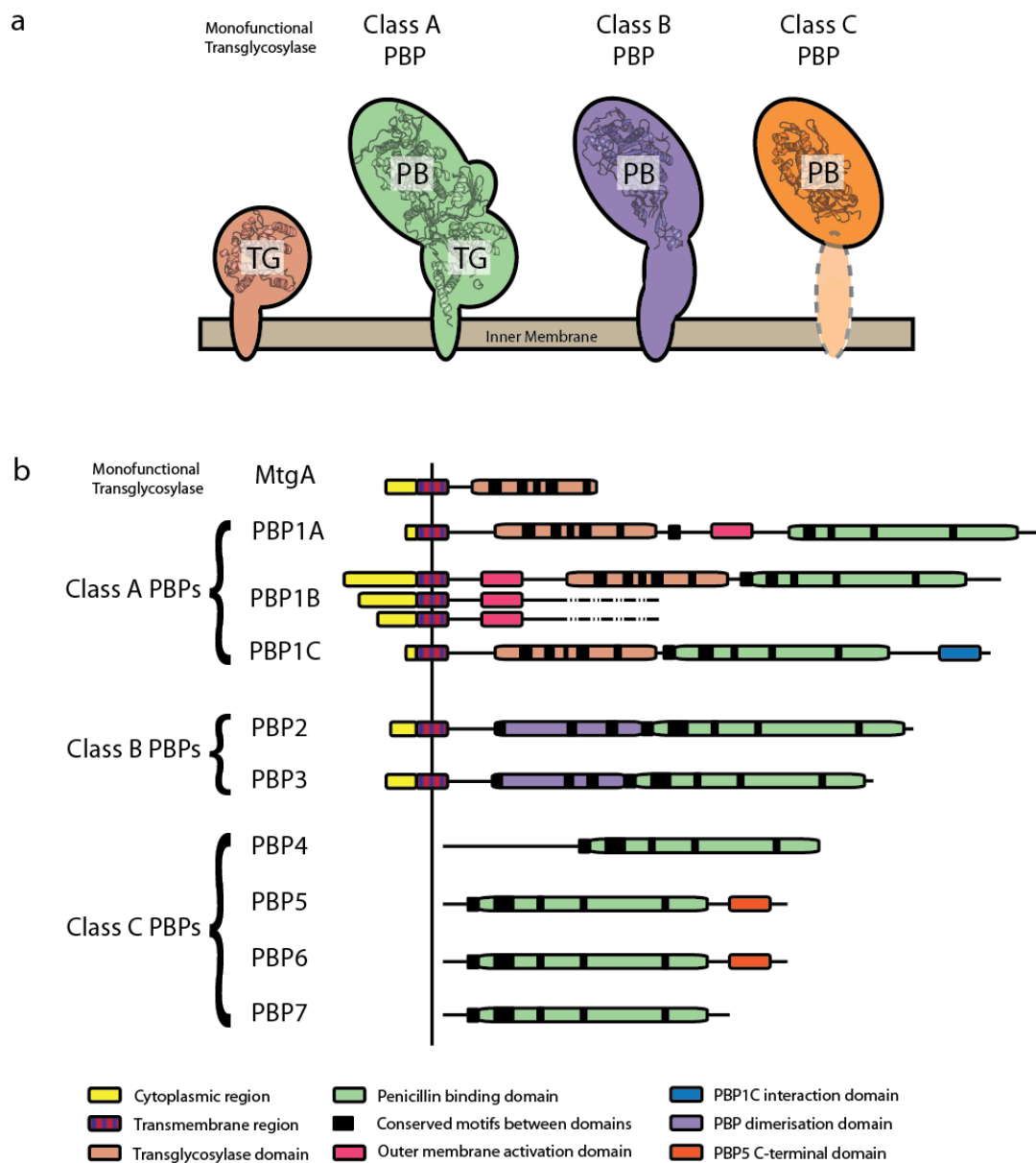


Figure 1-11 Penicillin binding proteins and their domain structure

(a) Orientation and domain composition of proteins involved in peptidoglycan synthesis. (b) Classes of penicillin binding protein and domains found in each type. Adapted from Vollmer, Bertsche 2008¹⁰⁸.

Penicillin binding protein 1B (Pbp1B) has three isoforms (α , β , and γ) which have varying lengths of the N-terminal cytoplasmic region (Figure 1-11)^{114,120–122}. Similar to Pbp1A an outer membrane activating protein, LpoB, is required for activation of Pbp1B enzyme function and a complex is formed with Pbp3 as well as Fts proteins, linking to the cytoplasmic cytoskeleton¹¹⁰. Pbp1B has both TGase and DD-TPase function (Figure 1-10).

Penicillin binding protein 1C (Pbp1C) has been shown to have TGase function and has the ability to bind moenomycin as do Pbp1A and Pbp1B¹²³. Moenomycin has structural similarity to peptidoglycan subunits and is shown to bind and inhibit TGases¹²³. Pbp1Cs TPase function is only inferred from the sequence identity to Pbp1A and Pbp1B (Figure 1-11)¹²³. Upon knockout of *pbpC*, the gene encoding Pbp1C, no notable change in growth or morphology is found and the presence of this gene is not able to compensate for the double knockout of *mrcA* and *mrcB* (encoding Pbp1A, Pbp1B, respectively)¹²³. Pbp1C has been shown to have lesser affinity to β -lactams than other PBPs with notable binding to the β -lactam antibiotics oxacephalosporin latamoxef and a Bolton/Hunter derivative of ampicillin¹²³. It is also said that as there is varied specificity to β -lactams and the lack of compensation of TPase function lost by the other Class A PBPs, that Pbp1C is no longer a functioning TPase enzyme¹²³. Pbp1C has been suggested to function as part of a complex as it can interact with Pbp1B, PBP2, PBP3 and MltA¹²³. Although LpoA and LpoB have been identified as essential for function of Pbp1A and Pbp1B respectively no equivalent has been identified for Pbp1C to date.

Class B penicillin binding proteins have 2 major domains with the N-terminal domain having some similarity to the TGase domain seen within Class A PBPs however without any direct TGase function^{84,107,124}. This N-terminal domain functions in the formation of peptidoglycan synthesis complexes with other PBPs and cytoskeletal elements. The second domain has DD-TPase function.

Pbp2 is an enzyme with DD-TPase activity essential for cell elongation in *E. coli* and for the characteristic rod shape seen (Figure 1-10, 1-11)^{117,118}. The gene encoding Pbp2, *pbpA*, is found within an operon alongside the gene *rodA*, which encodes RodA, a protein that

spans the inner-membrane and is also essential for extension of the cylindrical wall of the *E. coli* rod^{117,118}. When cells are grown with mecillinam, a Pbp2 specific β -lactam antibiotic, they form spheres instead of rods and are not able to form colonies^{117,118}. The exposure of bacteria to mecillinam to generate resistant mutants results in strains that can tolerate deletion of the gene encoding Pbp2^{117,118}. As this specific inhibition shows that Pbp2 is essential for cell elongation it has also been suggested as having an essential function in regulation of septation^{117,118}. When Pbp2 is tagged with green fluorescent protein it appears within the lateral cylindrical wall as well as at the site of cell division disappearing just before the cells divide^{117,118}. Cells that have been exposed to mecillinam are only able to divide for several generations before lysing which can be extended for further generations by overexpression of FtsZ which is involved in the division machinery of the cell^{117,118}. As a reduction of Pbp2 would result in spherical cells it is thought that upon inhibition of its function the demand for FtsZ is greater^{117,118}.

The monofunctional DD-TPase Pbp3 was once thought to have bifunctional TGase/TPase activity however the TGase like domain only has some of the conserved binding motifs found in TGase enzymes (Figure 1-11)^{113,125,126}. Pbp3 is encoded by the gene *ftsI* and is temperature sensitive, not functioning when grown at high temperatures. In the presence of the antibiotic aztreonam which has very high affinity for Pbp3, as well as at high temperatures that restrict enzyme function, cells grow in long filaments and are unable to divide¹²⁷. It is accepted that Pbp3 is an essential TPase involved in cell division that forms a complex with several proteins for peptidoglycan enlargement and cell division.

Monofunctional TGases although not penicillin binding proteins have a major role in the synthesis of the murein sacculus¹¹⁴. These enzymes catalyse the addition of lipid II peptidoglycan subunits to the generating peptidoglycan chain. Once the subunit has been added to the chain the diphosphorylated undecaprenol has a phosphate removed and the monophosphorylated undecaprenol returns to the cytoplasm for further transport of subunits. Monofunctional TGase enzymes like MtgA from *E. coli* are bound to the inner membrane and have five conserved motifs that are found within other enzymes with TGase function (Figure 1-11)¹¹⁴. Although the *mtgA* gene can be knocked out, this is

associated with reduced virulence within pathogenic bacterial strains and abnormal growth under laboratory conditions¹¹⁴.

1.2.3 Murein hydrolases

Murein hydrolases encompass all enzymes that cleave covalent bonds within the murein sacculus including muramidases that cleave between sugar subunits, peptidases that cleave between amino acid residues and amidases that cleave between sugar and peptide crosslinks^{128–132}.

Muramidase enzymes cleave between sugar subunits with lysozyme cleaving the β 1,4-bond between MurNAc and GlcNAc subunits in glycan chains and lytic transglycosylases (LTases) cleaving the glycosidic bond between them forming a concomitant 1,6-anhydro bond on the MurNAc residue^{128,129,133,134}. In *E. coli* five outermembrane lipoproteins (MltA, MltB, MltC, MltD, and EmtA) LTases have been identified in addition to one soluble LTase (Slt70)^{135–138}. *Endo*-LTases such as EmtA cleave sugars within the chain whereas *exo*-LTases (Slt70, MltA and MltB) cleave the disaccharide subunits from the 1,6-anhydroMurNAc glycan end of the chain (Figure 1-10). There are various types of peptidases associated with murein hydrolysis.

Carboxypeptidases (CPases) cleave amino acids from peptide side chains. Endopeptidases (EPases) hydrolyse the bridge formed between peptide chains by TPases^{139,140}. DD-CPases cleave between the D-alanine-D-alanine end residues of a pentapeptide sidechain and LD-CPases cleave between the meso-diaminopimelic acid and D-alanine (Figure 1-10). An important protein for recycling of peptidoglycan is LdcA which is an LD-CPase found within the cytoplasm. DD-EPases cleave the bridge formed between the meso-diaminopimelic acid and D-alanine of two side chains and LD-EPases cleave the bridge formed between two meso-diaminopimelic acids of two side chains (Figure 1-10). MepA is a protein 30 kDa in size that is found in the periplasm and has been shown to have both DD-EPase activity and LD-EPase activity¹⁴¹.

Class C penicillin binding proteins are peptidases with one main enzymatic domain that include both DD-CPases and DD-EPases. Pbp4 and Pbp7 are both located in the periplasm and have been shown to be membrane associated with DD-EPase activity; however Pbp4

has also been shown to have some DD-CPase activity (Figure 1-11)^{142–144}. Pbp5, Pbp6, and Pbp6B have been shown to have DD-CPase activity and although they do not have a membrane anchor, are thought to be associated with the inner-membrane (Figure 1-11)^{131,145}. These DD-CPases play an important role in cell shape with abnormalities seen upon combination of deletions between these and other enzymes involved in peptidoglycan synthesis and hydrolysis¹²⁷.

The third type of murein hydrolases, amidases, cleave the bond between glycan strand and peptide side chain (Figure 1-10). N-acetylmuramyl-L-alanine amidases AmiA, AmiB and AmiC all have an important role in cell separation during cell division^{130,146}. The fourth amidase identified, AmiD, is expressed in the cytoplasm and is involved in recycling of peptidoglycan subunits¹⁴⁷.

1.2.4 Murein structure

The murein sacculus is a large hetero-polymer structure found within the periplasm of Gram-negative bacteria that maintains the bacteria's shape and is essential to prevent rupture by the bacteria's internal turgor¹⁴⁸. This is made up of subunits consisting of a disaccharide of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) with a pentapeptide sidechain extending from the MurNAc sugar^{111,149} (Figure 1-12). This pentapeptide usually consists of a chain of L-alanine, D-glutamate, meso-diaminopimelic acid, followed by two D-alanine (Figure 1-12) with the D-glutamate residue linked via its γ -carboxyl group to the L-centre of meso-diaminopimelic acid^{150,151}. These subunits are synthesised within the cytoplasm and cross the cytoplasmic membrane to the periplasm in the form of lipid II and the peptidoglycan subunits are cross-linked into the existing peptidoglycan by penicillin binding proteins¹⁰⁷. The GlcNAc end of the subunit is cross-linked to the existing MurNAc chain by TGases and the peptide chains are then cross-linked by TPases, shortened by CPases or removed by amidases. The two ends of the glycan strands are either GlcNAc or 1,6-anhydroMurNAc, which is a modified MurNAc with a C-1, C-6 ether-linkage¹¹¹.

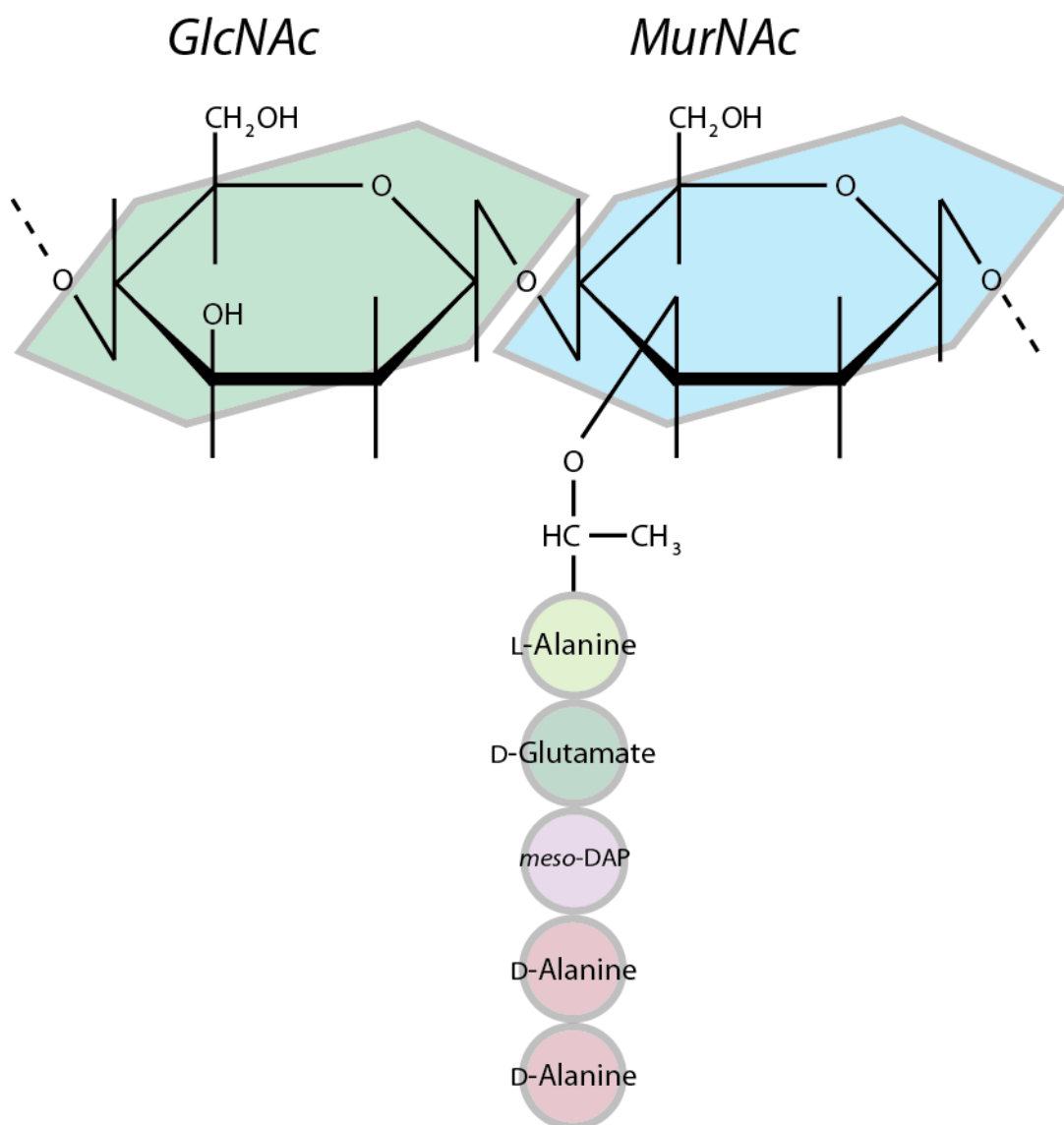


Figure 1-12 Schematic of the primary structure of peptidoglycan pentapeptide subunit.

Circles represent the amino acids labelled that are attached to MurNAc. GlcNAc, N-acetylglucosamine; *meso*-DAP, meso-diaminopimelic acid; MurNAc, N-acetylmuramic acid.

Several models have been suggested of how this network of glycan chains with peptide crosslinks forms the peptidoglycan network. Initial work gave evidence for glycan chains following a perpendicular angle to the direction of the rod shape with peptide links forming parallel to the rod shape ¹⁴⁸(Figure 1-13. a). The NMR structure for a dimer of the peptidoglycan subunit has been modelled with suggestion that the glycan chain is orientated to form chains between membranes forming a hexagon of peptide crosslinks which allows for the 120° angle of rotation measured between peptide crosslinks ¹⁵²(Figure 1-13. b, c). This fits well with the model of the outermembrane protein TolC that fits within this hexagonal network to span both the outermembrane and the peptidoglycan layer (Figure 1-13. d). At the same time as this work electron cryotomography was performed on the purified sacculus of gram-negative bacteria added evidence away towards a disordered circumferential layered model of peptidoglycan. The disordered circumferential layered model has glycan strands arranged perpendicular to lengthwise rod shape of the bacterium ¹⁵³. This perpendicular model of glycan strands to the rod shape of bacterium was further confirmed using atomic force microscopy (AFM) of whole cells of *Lactococcus lactis* ¹⁵⁴. The use of AFM and super resolution fluorescence microscopy on purified sacculus later suggested that the synthesis of new peptidoglycan is limited to areas with pores in the peptidoglycan allowing new growth ^{155,156}.

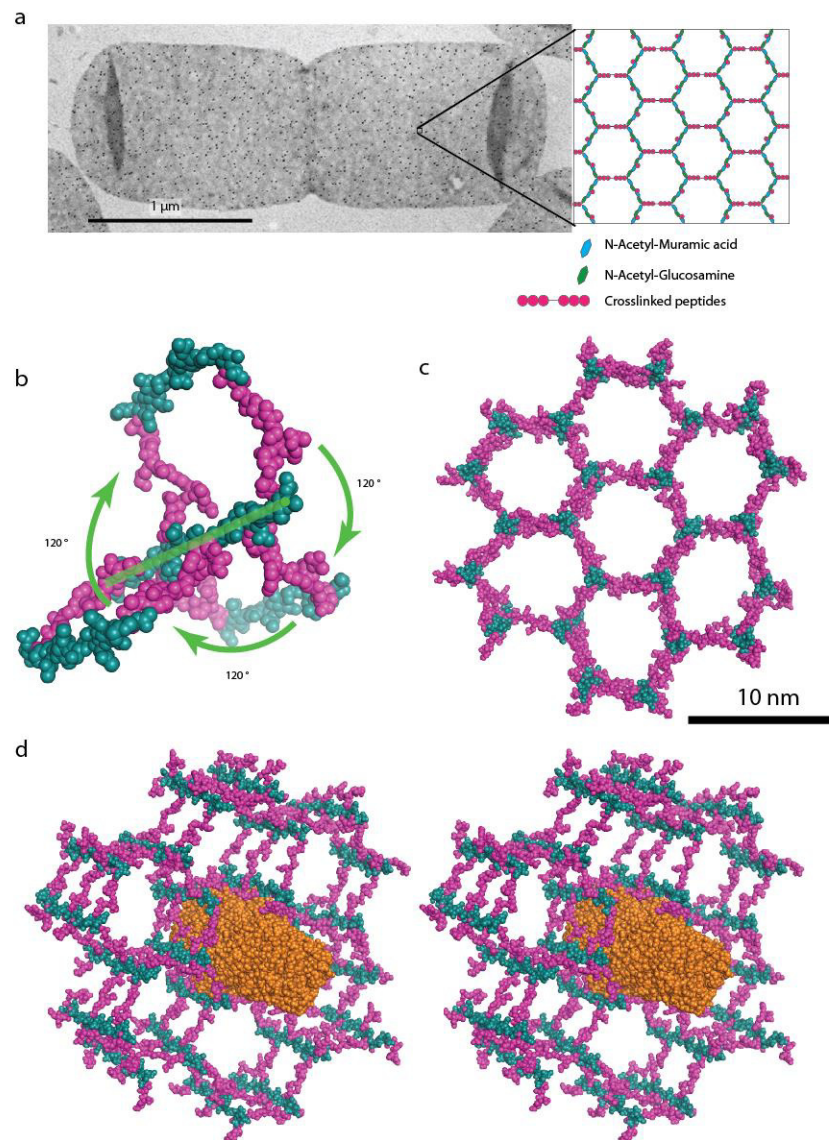


Figure 1-13 Electron micrograph of murein sacculus of *E. coli* and the three dimensional structure of peptidoglycan.

(a) Black dots within the electron micrograph show 6 nm gold particles bound to anti-murein antibody bound to the sacculus. Model on right shows an area approximately 30 x 30 nm. (b) Sphere representation of the three dimensional model of N-acetylglucosamine (GlcNAc), N-acetylmuramic acid (MurNAc) chain in teal and peptide sidechains in pink. With each peptidoglycan subunit there is a 120 rotation along the GlcNAc-MurNAc disaccharide axis between peptide crosslinks. (c) Top down view of model of sacculus. (d) Stereo view of peptidoglycan with TolC β -barrel domain in a hexagonal pore of the peptidoglycan, the protruding β -barrel would be bound in the outer membrane. Image in (a) was reproduced from Vollmer, Bertsche 2008¹⁰⁸. Structural model of murein used in (b), (c) and (d) provided on request by Shahriar Monbashery¹⁵².

1.3 *Aims of the project*

The overall aims of this project were to develop functional and structural understanding into ECAM and Pbp1C. These two proteins that are encoded together within an operon, have been described as functioning alongside each other, with ECAM inhibiting proteases and Pbp1C repairing peptidoglycan however we further set out to determine if there was any direct interaction between these proteins. At the outset of this work there had been limited biophysical characterisation of either ECAM or Pbp1C with only some minor analytical ultracentrifugation performed on ECAM. At the outset of this work a crystal structure of methylamine activated human α 2M had been published at 4.3 Å resolution. However, there had been no native α 2M structure published or an α 2M structure that had been activated through protease cleavage. Further it was hoped that through the use of knockouts and western blotting the behaviour of these proteins *in vivo* could be better understood.

2 Materials and Methods

2.1 Reagents

All chemicals and reagents used within this work were purchased from Melford, Sigma Aldrich or Thermo-Fisher Scientific unless stated otherwise.

2.2 Bacterial strains, plasmids and growth media

Bacterial strains and plasmids used in this work are shown in Table 2-1 and Tables 2-2, 2-3 respectively. Bacteria were grown in Lysogeny broth (LB; 10 g NaCl, 10 g tryptone, 5 g yeast extract per litre in dH₂O adjusted to pH 7.5) at 37°C. Bacteria were grown on agar plates and stored short term at 4°C and stored long term in 50% glycerol at -80°C.

Table 2-1 *E. coli* strains and knockouts

<i>E. coli</i> strain	Noted characteristics	Source
DH5α	<i>F</i> - Φ80 <i>lacZ</i> Δ <i>M15</i> , Δ(<i>lacZYA-argF</i>), <i>U169</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (<i>rk</i> -, <i>mk</i> +), <i>phoA</i> , <i>supE44</i> , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i> λ-	Invitrogen
BL21(DE3)	<i>fhuA2</i> [<i>lon</i>], <i>ompT gal</i> (λ <i>DE3</i>) [<i>dcm</i>], Δ <i>hsdS</i> , λ <i>DE3</i> = λ <i>sBamHlo</i> , Δ <i>EcoRI-B int</i> ::(<i>lacI</i> :: <i>PlacUV5</i> :: <i>T7 gene1</i>) <i>i21</i> , Δ <i>nin5</i>	New England Biolabs
T7 Express Crystal Competent <i>E. coli</i>	<i>fhuA2 lacZ</i> :: <i>T7 gene1</i> [<i>lon</i>] <i>ompT gal</i> <i>sulA11 R</i> (<i>mcr-73</i> :: <i>miniTn10</i> -- <i>TetS</i>)2 [<i>dcm</i>] <i>R</i> (<i>zgb-210</i> :: <i>Tn10</i> -- <i>TetS</i>) <i>endA1 metB1</i> Δ(<i>mcrC-mrr</i>)114::IS10	New England Biolabs
K-12 (BW25113)	<i>F</i> -, Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB-3</i>), λ-, <i>rph-1</i> , Δ(<i>rhaD-rhaB</i>)568, <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
K-12 mutants		
Δ <i>yfhM</i> <i>yfhM766</i> (del)::kan (JW2504-1)	<i>F</i> -, Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB-3</i>), λ-, Δ <i>yfhM766</i> ::kan, <i>rph-1</i> , Δ(<i>rhaD-rhaB</i>)568, <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
Δ <i>pbpC</i> <i>pbpC765</i> (del)::kan (JW2503-1)	<i>F</i> -, Δ(<i>araD-araB</i>)567, Δ <i>lacZ</i> 4787(:: <i>rrnB-3</i>), λ-, Δ <i>pbpC765</i> ::kan, <i>rph-1</i> , Δ(<i>rhaD-rhaB</i>)568, <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)

<i>ΔytfN</i> ytfN770(del)::kan (JW4180-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>rph-1, Δ(rhaD-rhaB)568, ΔytfN770::kan, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔytfM</i> ytfM769(del)::kan (JW4179-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>rph-1, Δ(rhaD-rhaB)568, ΔytfM769::kan, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔyfaA</i> yfaA784(del)::kan (JW2224-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔyfaA784::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔytaP</i> yfaP779(del)::kan (JW2219-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔyfaP779::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔytaQ</i> yfaQ780(del)::kan (JW2220-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔyfaQ780::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔytaS</i> yfaS781(del)::kan (JW2221-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔyfaS781::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔytaT</i> yfaT783(del)::kan (JW2223-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔyfaT783::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔmrcA</i> mrcA731(del)::kan (JW3359-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔmrcA731::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔmrcB</i> mrcB765(del)::kan (JW0145-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔmrcB765::kan, ΔlacZ4787(::rrnB-3), λ</i> -, <i>rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔpbpG</i> pbpG760(del)::kan (JW5355-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>ΔpbpG760::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔampC</i> ampC777(del)::kan (JW4111-2)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ</i> -, <i>rph-1, Δ(rhaD-rhaB)568, ΔampC777::kan, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔdacA</i> dacA771(del)::kan (JW0627-1)	<i>F</i> -, <i>Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), ΔdacA771::kan, λ</i> -, <i>rph-1, Δ(rhaD-rhaB)568, hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)

<i>ΔdacB</i> dacB789(del)::kan (JW3149-1)	<i>F</i> -, <i>Δ(araD-araB)567</i> , <i>ΔlacZ4787(::rrnB-3)</i> , <i>λ</i> -, <i>ΔdacB789::kan</i> , <i>rph-1</i> , <i>Δ(rhaD-rhaB)568</i> , <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔdacC</i> dacC739(del)::kan (JW0823-1)	<i>F</i> -, <i>Δ(araD-araB)567</i> , <i>ΔlacZ4787(::rrnB-3)</i> , <i>λ</i> -, <i>ΔdacC739::kan</i> , <i>rph-1</i> , <i>Δ(rhaD-rhaB)568</i> , <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>ΔopgH</i> opgH720(del)::kan (JW1037-2)	<i>F</i> -, <i>Δ(araD-araB)567</i> , <i>ΔlacZ4787(::rrnB-3)</i> , <i>λ</i> -, <i>ΔopgH720::kan</i> , <i>rph-1</i> , <i>Δ(rhaD-rhaB)568</i> , <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)
<i>Δprc</i> prc-755(del)::kan (JW1819-1)	<i>F</i> -, <i>Δ(araD-araB)567</i> , <i>ΔlacZ4787(::rrnB-3)</i> , <i>λ</i> -, <i>Δprc</i> - <i>755::kan</i> , <i>rph-1</i> , <i>Δ(rhaD-rhaB)568</i> , <i>hsdR514</i>	Coli Genetic Stock Center, Yale (Keio collection)

Table 2-2 Protein expression vectors

Plasmid	Noted Characteristics	Source
pET21-a(+)	Ampicillin ^r , cloning/expression vector, promoter; T7	Invitrogen
pET28-a(+)	Kanamycin ^r , cloning/expression vector, promoter; T7	Invitrogen
pBAD18-kan	Kanamycin ^r , cloning/expression vector, arabinose induced	Coli Genetic Stock Center, Yale
pKM54	pET21-a(+) with <i>yfhM</i> bases 48-4962 (ECAM residues 17-1653) inserted into NdeI/XhoI sites, N-terminal His ₆ tag	Dr Khedidja Mosbahi
pCF1 (ECAM)	pET21-a(+) with <i>yfhM</i> bases 55-4959 (ECAM residues 19-1653) inserted into NdeI/XhoI sites, C-terminal His ₆ tag, 6 mutations noted	This study
pCF2 (ECAM)	pET21-a(+) with <i>yfhM</i> bases 55-4959 (ECAM residues 19-1653) inserted into NdeI/XhoI sites, C-terminal His ₆ tag, 5 mutations noted	This study
pCF3 (ECAM)	pET21-a(+) with <i>yfhM</i> bases 55-4959 (ECAM residues 19-1653) inserted into NdeI/XhoI sites, C-terminal His ₆ tag, 4 mutations noted	This study
pCF4 (Pbp1C)	pET28-a(+) with <i>pbpC</i> bases 1-2313 (Pbp1C residues 1-770) inserted into XhoI/KpnI sites, N-terminal His ₆ tag	This study

pCF5 (ECAM)	pET21-a(+) with <i>yfhM</i> bases 55-4959 (ECAM residues 19-1653) inserted into NdeI/XhoI sites, 2 mutations noted with deletion resulting in no C-terminal His ₆ tag	This study
-------------	--	------------

Table 2-3 Complementation vectors

Plasmid	Noted Characteristics	Source
ppbpCc	pBAD18-kan with <i>pbpC</i> bases 1-2313 (Pbp1C residues 1-770) inserted into EcoRI/KpnI sites	This study
ppbpCo	pBAD18-kan with <i>pbpC</i> bases 1-2310 (Pbp1C residues 1-770) inserted into EcoRI/KpnI sites, N-terminal stop codon excluded	This study

2.3 Preparation of chemically competent cells

Magnesium chloride buffer (MCB; 0.1 M MgCl₂) and calcium chloride buffer (CCB; 0.05 M CaCl₂; 15% glycerol added post-autoclave) were autoclaved and chilled in iced water.

Strains of *E. coli* cells (50 ml) were grown to approximately 0.6 OD_{600 nm} in LB and pelleted at 4°C by centrifugation at 4000 g for 15 minutes. Cell pellets were washed and centrifuged twice using ice-chilled MCB before being resuspended in 2 ml CCB. Chemically competent cells were aliquoted in 200 µl fractions, frozen in liquid nitrogen and stored at -80°C.

2.4 Molecular biology

All restriction, DNA polymerase, and T4 DNA ligase enzymes were purchased from New England Biolabs (NEB).

2.4.1 Polymerase chain reaction (PCR)

PCR was used to amplify the targeted gene using the reaction protocol shown in Table 2-4 using components in Table 2-5 with PCR products electrophoresed on agarose gel.

Table 2-4 Amplification PCR reaction

Temperature (°C)	Step	Time (s)
95	Initial denaturing	30
95	Denaturing	45
52-60	Annealing	45
72	Extension	60/kb
~	Go to Denaturing step for 30 cycles	~
72	Final Extension	600
10	Hold	~

Table 2-5 Amplification PCR reaction components

Component	Stock Concentration	Volume (μl)
Phusion DNA polymerase	2 units/μl	1
Reaction buffer	5x	10
Reverse primer	1 nM	5
Forward primer	1 nM	5
Template	Variable	1-5
dNTPs	5 mM	1
dH ₂ O	100%	23-27
Total		50

Mutagenesis primers MutP1-P4 (Table 2-8) were used to perform PCR mutagenesis to correct four mutations in the ECAM clone pCF3 (Table 2-2). The four mutations were to base numbers 655, 1817, 4760 and 4782 resulting in t-c, a-g, g-a, t-c base changes respectively. The first three mutations resulted in amino acid changes of S-P, Q-R, S-N, respectively with the final mutation being silent. PCR mutagenesis was performed following the mutagenesis PCR reaction protocol shown in Table 2-6 using components in Table 2-7 with PCR products electrophoresed on agarose gel.

Table 2-6 Mutagenesis PCR reaction

Temperature (°C)	Step	Time (s)
95	Initial denaturing	30
95	Denaturing	30
55	Annealing	60
68	Extension	60/kb
~	Go to Denaturing step for 18 cycles	~
72	Final extension	600
10	Hold	~

Table 2-7 Mutagenesis PCR reaction components

Component	Stock Concentration	Volume (μl)
Phusion DNA polymerase	2 units/μl	1
Reaction buffer	5x	10
Reverse primer	1 nM	5
Forward primer	1 nM	5
Template	2.5 nM	1
dNTPs	5 mM	1
dH ₂ O	100%	27
Total		50

2.4.2 Oligonucleotide primers and sequencing

Primers used within this work were synthesised by Eurofins Operon¹⁵⁷. Sequencing was performed by Source BioScience¹⁵⁸.

Table 2-8 *yfhM* (ECAM) primers

Primer	Restriction enzyme site	Sequence 5' to 3' (restriction site bold)	Purpose
YfhMNdeIF	NdeI	TAGTAGTAC ATATG GACAACA ACGATAACGCGCCAAC	yfhM (ECAM) -54 bases for signal sequence, cloning forward
cfyhMns	XhoI	TTTGACGT CTCGAG CGGTCTGA CAATCAGCAGATC	yfhM (ECAM) -3 base stop codon, cloning reverse
YtfMCFseq1	n/a	ACGAAAAAACTATAACTACCCG CG	yfhM (ECAM) sequencing primer 1, forward
YtfMCFseq2	n/a	CCCAGCAAGGAATTGAAGTCTC	yfhM (ECAM) sequencing primer 2, forward
YtfMCFseq3	n/a	CGAAGATTTTATGCCAGAGCGC	yfhM (ECAM) sequencing primer 3, forward
YtfMCFseq4	n/a	CCGCTTTTGACATCGTTTATAG C	yfhM (ECAM) sequencing primer 4, forward
YtfMCFseq5	n/a	CCGGTCGATAAAACCTGGAATC	yfhM (ECAM) sequencing primer 5, forward
YtfMCFseq6	n/a	CAGGGCGAAGGCTCGGTTA	yfhM (ECAM) sequencing primer 6, forward
YtfMCFseq7	n/a	CCGGCGGATGGATTGCAAAAC	yfhM (ECAM) sequencing primer 7, forward
YtfMCFseq8	n/a	CGCTGGTGTGGCCCGTC	yfhM (ECAM) sequencing primer 8, forward
MutP1aF	n/a	CAAAGTCTGGAAAACGGCGCG GCCAG	ECAM mutation repair mutagenesis forward 1
MutP1bR	n/a	CTGGGCCGCGCCGTTTCCAGA CTTTG	ECAM mutation repair mutagenesis reverse 1
MutP2aF	n/a	GCCCAGCAAGGAATTGAAGTC TCTTTATTAATGAGAAAGGG	ECAM mutation repair mutagenesis forward 2
MutP2bR	n/a	CCCTTTCTCATTTAATAAAGAG ACTTCAATTCCTTGCTGGGC	ECAM mutation repair mutagenesis reverse 2
MutP3aF	n/a	GTAGTGTTCTGTTTCTGGGCTGG CTATAGC	ECAM mutation repair mutagenesis forward 3

MutP3bR	n/a	GCTATAGCCAGCCCAGAAACG AACACTAC	ECAM mutation repair mutagenesis reverse 3
MutP4aF	n/a	GATGGAGAAATTCAGGCCACC ATTAGCGGG	ECAM mutation repair mutagenesis forward 4
MutP4bR	n/a	CCCGCTAATGGTGGCCTGAATT TCTCCATC	ECAM mutation repair mutagenesis reverse 4

Table 2-9 *pbpC* (Pbp1C) primers

Primer	Restriction enzyme site	Sequence 5' to 3' (restriction site bold)	Purpose
CFPbpCNde1-105F	NdeI	GCAAC CATATG CTGCATGAAGT CAATCCCGC	pbpC (Pbp1C) -105 base, cloning forward
CFPbpCNde1-135F	NdeI	GCAAC CATATG GTGGCGCAGG ATGGTACG	pbpC (Pbp1C) -135 base, cloning forward
CFPbpCNde1-165F	NdeI	GCAAC CATATG TTCGCCGATGC TGACGGC	pbpC (Pbp1C) -165 base, cloning forward
CFPbpCXho1R	XhoI	TACTA CTCGAG CTATTGCATGA CAAATTTCACTGTC	pbpC (Pbp1C) cloning reverse
cfpbpCssXho1F	XhoI	TAGTAG CTCGAG ATGCCTCGCT TGTTAACCAAACG	pbpC (Pbp1C) cloning forward
cfpbpCssKpn1R	KpnI	CATA CTGGTAC CTATTGCATG ACAAATTTCACTGTCG	pbpC (Pbp1C) cloning reverse
pbpCEcoR1F	EcoRI	GCTGCT GAATTC ATGCCTCGCT TGTTAACCAAACG	pbpC (Pbp1C) cloning forward, complementation vector
pbpCKpn1RA	KpnI	GCTGCT GGTAC CTATTGCATG ACAAATTTCACTGTCGCG	pbpC (Pbp1C) cloning reverse, complementation vector closed
pbpCKpn1RB	KpnI	GCTGCT GGTAC CTTGCATGACA AATTTCACTGTCGCGA	pbpC (Pbp1C) -3 base stop codon, cloning reverse, complementation vector open for including myc tag

2.4.3 DNA restriction enzyme digest

Purified DNA (10-20 µg) was added to 20 units of appropriate restriction enzymes and 1 µl of 10x Cutsmart buffer (NEB, UK) and made to a final volume of 15 µl. Restriction

digests were incubated at 37°C for 3-4 hours and separated by agarose gel electrophoresis.

2.4.4 Agarose gel electrophoresis

Agarose was melted by microwave at 0.8% (w/v) agarose in TBE buffer (0.17 M Tris, 0.2 M borate, 5 mM EDTA, pH 8). The melted agarose was cooled to approximately 50°C and GelRed (Cambridge Bioscience, UK) was added (1:10000) and poured into the electrophoresis apparatus. The set gel was submerged in TBE buffer and samples mixed with 5x DNA loading dye (Bioline, UK) were loaded onto the gel alongside a 1 kb+ DNA ladder (Invitrogen) as a base pair size reference. The gel was run for 40-60 minutes at 60 V and visualised using a UVipro Gold transilluminator (UVIttec, UK).

2.4.5 Agarose gel DNA extraction

Extraction of DNA from agarose was performed using a Qiagen gel extraction kit (Qiagen, UK) following the manufacturers protocol. DNA was eluted using elution buffer (10 mM Tris, pH 8.0) and stored at -20°C.

2.4.6 DNA ligation

A ligation reaction containing 3:1 ratio of digested PCR insert to digested plasmid alongside 5 Units of T4 DNA ligase and 1X T4 DNA ligase buffer with a total final volume of 15 µl. DNA ligation reaction was incubated overnight at room temperature with 5 µl being used for transformation into chemically competent DH5α cells.

2.5 Transformation of chemically competent *E. coli* cells

Chemically competent *E. coli* cells (50-100 µl) were thawed on ice with 50-100 ng of plasmid. DNA ligation transformations consisted of 5 µl of ligation reaction added to 50-

100 µl of chemically competent cells. Cells were incubated on ice for 30 minutes before being heat shocked at 42°C for 45 seconds. 500 µl of LB was added and cells were allowed to recover at 37°C for 60 minutes and were then plated out onto LB agar plates containing appropriate antibiotics and grown overnight at 37°C.

2.6 Protein overexpression and purification

2.6.1 Cloning expression and purification of ECAM

The gene for ECAM, *yfhM* from *E. coli* K-12, was amplified by PCR using primers YfhMNdeIF and cfyfhMns, and cloned into pET21a vector using NdeI and XhoI restriction sites. The first 22 residues from the N-terminus of the gene, containing the signal sequence identified using SignalP, were excluded from the construct. The stop codon was also excluded, resulting in a protein containing residues 23-1631 and a C-terminal 6xHis-tag (LEHHHHHH). The first plasmid generated pCF1, was sequenced using primers YtfMCFseq1-8 and found to have 6 base mutations resulting in 6 amino acid changes. The second plasmid cloned, pCF2, was found to have 5 mutations and the third plasmid, pCF3 contained 4 mutations. ECAM plasmid pCF3 was mutated using MutP1a/b-4a/b however only the MutP1a/b primers were successful and resulted in plasmid pCF5 which contained two mutations and a deletion resulting in no 6xHis-tag. As the three amino acid residue changes identified in pCF3 (detailed on page 39) were not expected to cause change in protein function it was decided to use pCF3 for all future expression of ECAM. ECAM was expressed in *E. coli* BL21 (DE3) grown in LB to 0.6 OD_{600nm} and induced using 1 mM D-isopropyl-β-thiogalactopyranoside (IPTG) with cells grown for a further 6 hours. The cell pellet was collected by centrifugation at 4,400 g for 15 minutes and cells were re-suspended in binding buffer (20 mM Tris, 10 mM imidazole, 500 mM sodium chloride, pH 7.5) and lysed by sonication with 2 mg ml⁻¹ lysozyme in the presence of protease inhibitors (Complete Mini, Roche). Cell debris was removed by centrifugation at 46,000 g for 30 minutes at 4°C. Cell supernatant was then loaded onto a HisTrap™ HP column (GE Healthcare) and bound protein was eluted with elution buffer (20 mM Tris, 500 mM imidazole, 500 mM NaCl, pH 7.5) using a linear gradient increasing from 10 mM to 500 mM. Fractions containing ECAM were pooled and dialysed overnight at 4°C into 50 mM

Tris, 200 mM NaCl, pH 7.5 and run on a Superdex S200 gel-filtration column (GE Healthcare). Central fractions from the peak were combined and concentrated using a 100 kDa molecular weight cut off centrifugal concentrator.

ECAM was later nickel affinity purified in the presence of TCEP using a binding buffer and elution buffer supplemented with 5 mM TCEP (tris(2-carboxyethyl)phosphine) that was dialysed out prior to SEC.

ECAM was expressed for crystallography in T7 Express Crystal Competent *E. coli* (methionine auxotrophic strain, New England Biosciences), using an inducible T7 promoter and 1 mM IPTG as the inducer. Bacteria expressing selenomethionine-labelled ECAM were grown using M9 minimal media (35 mM Na_2HPO_4 , 22 mM KH_2PO_4 , 8.5 mM NaCl, 18.7 mM NH_4Cl , 2 mM MgSO_4 , 0.2 mM CaCl_2 , 0.2% D-glucose (anhydrous)) supplemented with 50 mg L^{-1} selenomethionine and 20 mg L^{-1} of each of nine essential amino acids (excluding methionine). Cells were grown at 37°C to an OD_{600} of 0.6 and protein production was induced by addition of 1 mM IPTG and cells were grown for a further 6 hours.

2.6.2 Cloning expression and purification of Pbp1C

Full length Pbp1C was initially expressed using plasmid pKM54 that uses ampicillin as a selection antibiotic. Penicillin binding proteins are inactivated by β -lactam antibiotics such as ampicillin and so *pbpC* was re-cloned into pET28a as this plasmid uses kanamycin as a selection antibiotic. Primers cfpbpC_{ssXho1F} and cfpbpC_{ssKpn1R} were used to amplify full length *pbpC* and was transformed into pET28a generating plasmid pCF4. Pbp1C was expressed in *E. coli* BL21 (DE3) grown in LB to 0.6 $\text{OD}_{600\text{nm}}$ and induced using 1 mM IPTG with cells grown for a further 6 hours. The cell pellet was collected by centrifugation at 4,400 g for 15 minutes and cells were re-suspended in binding buffer (20 mM Tris, 10 mM imidazole, 500 mM sodium chloride, pH 7.5) and lysed by sonication with 1 mg ml^{-1} lysozyme in the presence of protease inhibitors (Complete Mini, Roche). Cell debris was removed by centrifugation at 46,000 g for 30 minutes at 4°C. As Pbp1C expression

resulted in inclusion bodies without any soluble protein a protocol was followed to solubilise, refold and purify.

2.6.2.1 Solubilising Pbp1C

The inclusion body pellet containing Pbp1C was first suspended in 40 ml of wash buffer 1 (50 mM Tris, 200 mM NaCl, pH 7.5 with 0.1% lauryldimethylamine N-oxide (LDAO)) before being centrifuged at 4400 g to form a pellet. The pellet was washed twice using wash buffer 1 and once with wash buffer 2 (50 mM Tris, 200 mM NaCl, pH 7.5). For every 750 mg of washed wet pellet, 100 ml of Solubilisation Buffer (50 mM Tris, 200 mM NaCl, pH 7.5, 5% Sarkosyl) was added and resuspended until the pellet was soluble and no pellet could form through centrifugation.

2.6.2.2 Refolding and purification of Pbp1C

Solubilised Pbp1C was refolded by dialysing overnight at 4°C into Refolding Buffer (0.4% LDAO, 0.1 M L-arginine, 50 mM Tris, 200 mM NaCl, pH 7.5) and subsequently dialysed into SEC buffer (200 mM NaCl, 50 mM Tris, 0.1% LDAO, pH 7.5) for purification by size exclusion chromatography. Size Exclusion Chromatography (SEC) was performed on an ÄKTA purifier. Pbp1C samples to be purified were loaded onto S200 SEC Superdex column with Pbp1C SEC buffer (200 mM NaCl, 50 mM Tris, 0.1% LDAO, pH 7.5).

2.6.3 Protein concentration

ECAM and Pbp1C were concentrated using Vivaspin 20 (Sartorius) concentrator tube (100 kDa molecular weight cut off) by centrifugation at 4400 g.

2.7 Protein concentration determination

The ExPASy online program ProtParam was used to calculate the molar extinction coefficient of proteins^{159,160}. Protein samples then had their absorbance measured at 280

nm wavelength over a path length of 1 cm, and these values were applied to the formula below to calculate protein concentration.

$$C = \frac{A}{\epsilon}$$

C = Protein concentration

A = Absorbance of sample measured at 280 nm wavelength over a path length of 1 cm

ϵ = Molar extinction coefficient

2.8 SDS-PAGE

Bacterial pellets or protein samples were mixed with Laemmli buffer (200 mM Tris, pH 6.8, 8% (v/v) SDS, 0.4% (v/v) bromophenol blue, 40% (v/v) glycerol, 4.7% (v/v) β -mercaptoethanol) at a 4:1 ratio, heated at 95°C for 5-10 minutes and centrifuged (500 g) for approximately 1 minute and loaded onto the gel alongside a broad range molecular weight marker (2-212 kDa (NEB, UK) or 3-260 kDa (Novex prestained marker, Thermo Fisher Scientific)). SDS-PAGE was performed using tris-glycine running buffer (0.02 M Tris, 0.25 M glycine, 5% (w/v) SDS) at 30 mA until the buffer front reached the bottom of the gel (approximately 45-60 minutes).

For samples to be analysed by mass spectrometry analysis of protease cleaved ECAM, crystals were washed in reservoir solution before being dissolved in deionised water and heated to 95°C in Laemmli buffer for 5 minutes. These were then were run on NuPAGE® Novex® 4-12% Bis-Tris gel (Invitrogen) and visible bands were cut for proteomic analysis.

2.9 Western blotting

For immunoblotting samples, bacterial pellets that had been run by SDS-PAGE were transferred onto nitrocellulose membranes using an ECL semi-dry transfer unit at 7 mA overnight. Transferred membranes were blocked using blocking buffer (20 mM Tris, 200 mM NaCl, 0.01% (v/v) Tween 20, 5% (w/v) skimmed milk (Marvel, UK) on a shaker for an hour at room temperature. Blocking buffer was poured off and rabbit anti-ECAM primary

antibody was diluted 1 in 3000 with Wash buffer (20 mM Tris, 200 mM NaCl, pH 7.2, 0.01% Tween 20) and added to the nitrocellulose and placed on a shaker for 1 hour. The nitrocellulose membrane was then washed 3 times for 20 minutes while on shaker using wash buffer. Goat anti-rabbit secondary antibody coupled with horseradish peroxidase (HRP) was then diluted 1 in 20,000 and incubated with the nitrocellulose membrane for 30 minutes on a shaker at room temperature. The nitrocellulose membrane was then washed three times 20 minutes while on shaker at room temperature. The Western blot was developed using enzyme-linked chemiluminescence Western BrightTM Sirius chemiluminescent substrate (Advansta, USA) on film.

2.10 Mass spectrometry

Peptide fragments from trypsinated excised bands were separated by liquid chromatography and subsequently had their mass to charge ratio measured by Orbitrap XL Mass spectrometry. These peptide fragments were then compared to a database of *in silico* digested proteins identifying proteins and fragments within excised bands. Samples were digested by trypsin and analysed by LC-MSMS (Orbitrap XL), performed at the Fingerprints Proteomics Facility at the University of Dundee¹⁶¹.

2.11 Circular dichroism (CD) spectroscopy

Far-UV CD was carried out on Pbp1C (6 μ M) in 0.1 M sodium phosphate pH 7.5 with 0.1% (v/v) LDAO using a Jasco J-810 spectropolarimeter. Spectra were collected in 0.02 cm path length quartz cuvettes. Data were analysed with the CONTIN procedure that is available from the Dichroweb server^{162,163}.

2.12 Reaction of ECAM with proteases

Purified ECAM was reacted at a 1:1 molar ratio with porcine elastase (MP Biomedicals) in 50 mM Tris, 200 mM NaCl, pH 7.5 on ice for 5 minutes before being loaded onto a Superdex S200 gel-filtration column (GE Healthcare). The two major peaks from gel

filtration were concentrated to 16 mg ml^{-1} separately using a 100 kDa molecular weight cut off centrifugal concentrators and used in crystallization trials.

Purified ECAM was reacted at a 1:1 molar ratio with chymotrypsin (Sigma-Aldrich) in 50 mM Tris, 200 mM NaCl, pH 7.5 on ice for 5 minutes before being loaded onto a Superdex S200 gel-filtration column (GE Healthcare). The major peak from gel filtration was concentrated to 4.0 mg ml^{-1} using a 100 kDa molecular weight cut off centrifugal concentrator for AUC and SAXS experiments.

2.13 Reaction of ECAM with methylamine

Purified ECAM was reacted with a final concentration of 5 mM methylamine (Sigma-Aldrich) in 50 mM Tris, 200 mM NaCl, pH 7.5 on ice for 5 minutes before being loaded onto a Superdex S200 gel-filtration column (GE Healthcare). The peak from gel filtration was concentrated to 4.6 mg ml^{-1} using a 100 kDa molecular weight cut off centrifugal concentrator for AUC and SAXS experiments.

2.14 Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were carried out in a Beckman Coulter Optima XL-I analytical ultracentrifuge using an An-50 Ti eight-hole rotor. ECAM concentrations ranged from 0.078 mg ml^{-1} to 5.0 mg ml^{-1} ECAM (purified with TCEP) concentrations ranged from 0.2 mg ml^{-1} to 6.0 mg ml^{-1} ECAM (with TCEP) concentrations ranged from 0.09 mg ml^{-1} to 5.4 mg ml^{-1} ECAM reacted with methylamine concentrations ranged from 0.073 mg ml^{-1} to 4.6 mg ml^{-1} and ECAM reacted with chymotrypsin concentrations ranged from 0.06 mg ml^{-1} to 4.0 mg ml^{-1} . Samples were loaded into 12 mm path-length charcoal-filled epon double-sector centrepieces (360 μl per sample) and the highest concentration sample was loaded into a 3 mm path-length centrepiece (90 μl) and spun at 49,000 rpm for ~ 9 h at 4°C . The outer sector of centrepieces had equivalent volume of buffer that the protein samples were dialysed against previously. Scans were collected every 7 minutes using both interference and absorbance optics (280 nm; a radial range of 5.8 - 7.2 cm, and

radial step-size of 0.005 cm were used). SEDNTERP¹⁶⁴ was used to calculate the partial specific volume (from the amino acid sequence of ECAM), the buffer density and viscosity at 20°C and 4°C (Table 2-10).

Table 2-10 ECAM partial specific volumes, buffer densities and viscosities calculated using SEDNTERP for the analysis of SV data

Buffer	Partial specific volume (ml/g)	Buffer density (g/ml)	Buffer viscosity (P)
50 mM Tris, 200 mM NaCl, pH 7.5	0.728 (4°C)	1.010 (4°C)	0.0162 (4°C)
	0.735 (20°C)	1.008 (20°C)	0.0104 (20°C)
50 mM Tris, 200 mM NaCl, 5 mM TCEP, pH 7.5	0.728 (4°C)	1.010 (4°C)	0.0162 (4°C)
	0.735 (20°C)	1.008 (20°C)	0.0104(20°C)

Data were analysed using SEDFIT to obtain the apparent sedimentation coefficients of ECAM using the continuous $c(s)$ distribution model¹⁶⁵. The apparent sedimentation coefficients were then used to calculate the concentration-independent $s_{20,w}^0$ by plotting sample versus apparent $s_{20,w}$ and extrapolating the graph to infinite dilution. Ultrascan SOMO was used to compute theoretical sedimentation coefficients from high-resolution X-ray structures¹⁶⁶.

2.15 Protein crystallisation and data collection

Several hundred crystallization conditions were tested including JCSG-plus, MIDAS™ and Morpheus® screens (Molecular Dimensions) for both concentrated peaks from reaction of ECAM with porcine elastase. A Cartesian Honeybee 8 + 1 (Harvard Bioscience) robot was used with 96 well plates dropping 0.5 µl of reservoir solution with 0.5 µl of protein sample. Subsequent scaled up crystal growth was performed using 2.5 µl of reservoir solution with 2.5 µl of protein sample. The initial crystal was grown in conditions containing 0.1 M potassium chloride, 0.1 M HEPES, 25% SOKALAN® CP 7, pH 7.0 and upon

optimisation the pH was adjusted to 7.5 for larger crystal growth. Crystals were grown using equal volumes of porcine elastase cleaved ECAM and reservoir solution using sitting drop vapour diffusion with crystals appearing after 2 days at 16°C for the second fraction and after 2 weeks for the first fraction. Cryoprotection was optimised with a 3:2 ratio of xylitol saturated reservoir solution to reservoir solution. Crystals were briefly soaked and flash cooled in liquid nitrogen for data collection. The best diffraction obtained was 3.8 Å, and molecular replacement with methylamine activated α 2M (PDB:4ACQ) was unsuccessful most likely as sequence identity with the human homologue was low (12%) and due to the difference in domain orientation between structural models¹³. Further expression was performed using a methionine auxotrophic strain of *E. coli* BL21 (T7 Express Crystal Competent *E. coli*, New England Biolabs) and purification and crystallization screens were repeated using selenomethionine labelled protein. As repeating the previous protocol with selenomethionine protein was unsuccessful, in situ proteolytic cleavage screening was performed using porcine elastase. Successful crystallization was achieved using a 1:100 ratio of porcine elastase to selenomethionine labelled ECAM. Crystallization was successful in the same condition as used previously with the crystal having a similar appearance and the same space group as previous unlabelled crystals. These crystals diffracted to 3.65 Å and phases were obtained using single-wavelength anomalous diffraction (SAD).

Data were collected for ECAM crystals at the I02, I03 and I24 beamlines at Diamond Light Source (Didcot, United Kingdom) at 100 K at the selenium K-edge ($\lambda = 0.97939$ Å) using a Pilatus 6M detector. A high redundancy SAD data set was processed and scaled using XDS and AIMLESS from the CCP4 suite of programs^{167–170}. Selenium sites were located using *Shelx C/D* with the best substructure solution consisting of 19 sites although the full length ECAM structure was predicted to have 23 selenomethionines¹⁷¹. These selenium sites were input with the SAD dataset to *Autosol* within the *Phenix* package to perform phasing and electron density modification¹⁷². Density modification in Autosol was sufficient to break phase ambiguity, due to the high solvent content of the crystal (69%)¹⁷². This yielded interpretable, low resolution, maps in which density corresponding to secondary structure elements and larger amino acid sidechains were visible. Initially,

the α -helical TED was built in COOT using idealised α -helical sections¹⁷³. The loops between these sections were connected where density was visible. The 6 selenomethionine sites in this domain provided the start sites for the building of amino acid side chains in the whole TED, although due to the relatively low resolution, initially only larger side chains and those where continuous sequence could be determined were built. In addition to the building of the TED, a number of α 2M derived poly-alanine MG domains were rigid body fitted into their corresponding density and manually real-space refined in COOT¹⁷³. As with the TED, where possible, side chains were fitted using the positions of selenomethionine Se atoms as start sites. This initial building yielded a partial model which was then used, in conjunction with the selenomethionine substructure to rephase experimental data using MR-SAD phasing in *Phaser* EP¹⁶⁹. This process led to phase improvement and the appearance of new features in the map, which were modelled and the process was repeated iteratively. Partway through the building process the atomic coordinates for SaA2M (which shares 82% sequence identity with ECAM) were published and the domains from this model provided validation for MG domain placement and sidechain modelling in our experimentally phased map (PDB: 4U48¹⁴). The SaA2M model also provided a template for building the more difficult sections of the model. At this point restrained TLS refinement using *REFMAC5* was found to stably improve both R_{work} and R_{free} . Such refinement was performed and the model improved and was finished manually in *COOT*^{173,174}. Electron density for the thioester bond indicated that the deaminated glutamine has no covalent bond with the cysteine. Before submission of the final model, quality of the structure was assessed using the Molprobity webserver¹⁷⁵. The atomic coordinates and structure factors were deposited in the Protein Data Bank (PDB: 4RTD).

2.16 Small angle X-ray scattering (SAXS)

SAXS was carried out on the X33 beamline at the Deutsches Elektronen Synchrotron (DESY, Hamburg, Germany). Data were collected on samples of ECAM and ECAM reacted with chymotrypsin at concentrations in the range of 0.4-6 mg ml⁻¹. Scattering curves for buffer were acquired before and after each sample and an average of the buffer

scattering was subtracted from the sample scattering. The data obtained for each sample were analysed using PRIMUS, merging scattering data for low concentration samples at low angles with high angle data for high concentration samples¹⁷⁶. The distance distribution function, $p(r)$, was obtained by indirect Fourier transform of the scattering intensity using GNOM¹⁷⁶. A Guinier plot ($\ln I(s)$ vs s^2) was used to determine the molecular weight and radius of gyration, R_g , of ECAM. *Ab initio* models of the protein in solution were built using DAMMIF¹⁷⁷⁻¹⁷⁹. DAMMIF models were averaged using DAMAVER and filtered using DAMFILT^{177,179}. CRY SOL was used to compute theoretical scattering curves from high-resolution X-ray structures¹⁸⁰.

3 Overexpression, purification and characterisation of ECAM

Small angle X-ray scattering data was collected by Inokentij Josts.

3.1 *Introduction*

It has been suggested that ECAM and PbpC function as a defence and repair mechanism against host proteases, with ECAM functioning as a protease inhibitor and PbpC repairing proteolytic damage to the peptidoglycan layer⁶. Among the various eukaryotic α 2Ms, some have been shown to form tetramers with others forming dimers and monomers^{16,17,21}. In contrast, ECAM is monomeric prior to protease activation, although once cleaved exists as a mixture of monomer, dimers and trimers in solution, with the formation of these multimeric forms being concentration dependent⁶⁷. In addition reaction of ECAM with methylamine does not cause a change in conformation as seen for α 2M⁶⁷. The high resolution X-ray crystal structures of native and methylamine-activated SaA2M show that methylamines only effect on SaA2M is localised to the thioester region of the TED¹⁴. This high resolution model of SaA2M was suggested as being a monomer however a crystallographic dimer was also present within its crystal lattice¹⁴. In order to begin structurally characterising ECAM sedimentation velocity (SV) experiments were performed to determine the oligomeric state of the protein in solution. SAXS was further used to characterise ECAM in solution.

3.2 Results

3.2.1 Expression and purification of ECAM

The gene encoding ECAM, *yfhM*, was cloned minus the region encoding the signal sequence and including a C-terminal 6-histidine tag. The gene amplified by PCR was cloned into pET21a using NdeI and XhoI. This plasmid was transformed into BL21(DE3) *Escherichia coli* cells for inducible expression. Cells were grown at 37°C to an OD₆₀₀ of 0.6 and 0.1 mM IPTG was added to induce ECAM expression. Cells were grown for a further four hours, lysed, and purified by Ni²⁺-affinity chromatography (Figure 3-1 a). Fractions containing purified ECAM were pooled and dialysed against 50 mM Tris-HCl, 200 mM NaCl, pH 7.5 and further purified on a Superdex S200 gel-filtration column (Figure 3-1 b, c). ECAM eluted as two peaks, with the first eluting at 147 ml and the second at 169 ml with the second peak approximately 8 times larger (Figure 3-1 c). The presence of two peaks on SEC was the first indication of potential oligomerisation in solution. When both peaks were compared by SDS-PAGE at the same concentration they have an identical composition with major bands at 260 kDa and 160 kDa (Figure 3-1 c). This suggests that there is a monomer and a dimer with a dimer being visible after SDS PAGE due to an interaction that survives heating and denaturation. Fractions from the main peak (169 ml) were collected and pooled for further experiments (Figure 3-1 c).

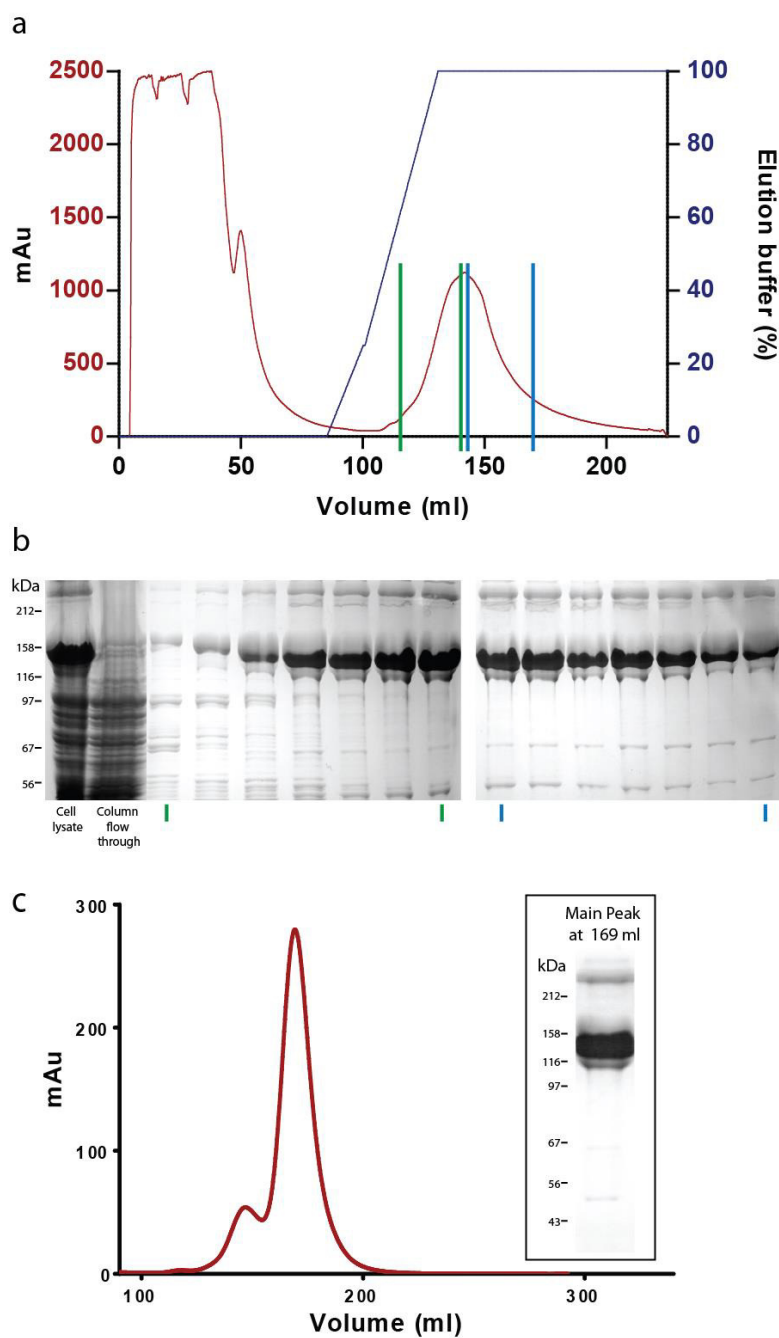


Figure 3-1 Purification of ECAM by Ni^{2+} -affinity chromatography and size exclusion chromatography (SEC) with SDS PAGE.

(a) An elution buffer (blue) gradient eluting ECAM (red) from a Ni^{2+} -affinity column. (b) Highlighted fractions corresponding to the Ni^{2+} -affinity elution fractions seen in (a). (c) Size exclusion chromatography of Ni^{2+} -affinity purified ECAM with inset image of main pooled fraction of purified protein around 169 ml volume.

Published purification protocols for ECAM in general include reducing agents such as DTT in buffers during cell lysis and Ni²⁺-affinity purification^{67,77}. To determine the effect of the inclusion of a reducing agent on the purification of ECAM, cells expressing ECAM were lysed in buffer containing 5 mM TCEP and purified by Ni²⁺-affinity chromatography (Figure 3-1 a, 3-2 a). ECAM was then dialysed to remove TCEP and imidazole before further purification by SEC where ECAM eluted as a single peak with an elution volume of 152 ml (Figure 3-2 b). This is considerably earlier than the major peak from SEC of ECAM purified in the absence of TCEP (169 ml) but is closer to the elution volume of the presumed dimeric species (147 ml) (Figure 3-2 b). The apparent difference in the oligomeric state of ECAM under different conditions was further investigated by SV.

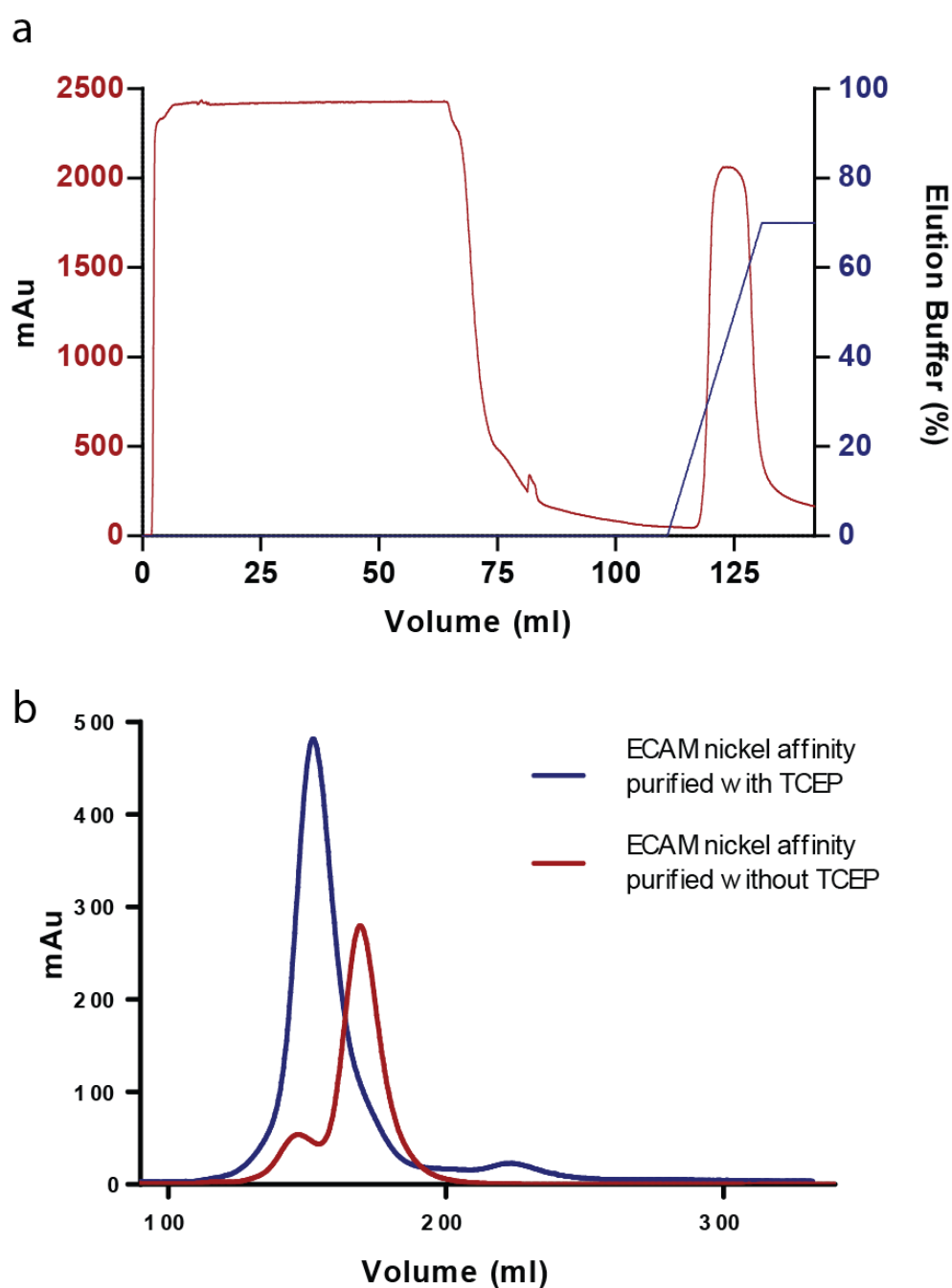


Figure 3-2 Purification of ECAM in the presence of TCEP by Ni²⁺-affinity chromatography and size exclusion chromatography (SEC) with SDS PAGE.

(a) An elution buffer (blue) gradient eluting ECAM (red) in the presence of 5 mM TCEP from a Ni²⁺-affinity column. (b) SEC of ECAM that had been Ni²⁺-affinity purified in the presence of 5 mM TCEP (blue) compared to ECAM without TCEP (red).

3.2.2 ECAM is in a monomer-dimer equilibrium in solution

Sedimentation velocity (SV) enables the oligomeric state of proteins to be determined as well as providing information on protein conformation. ECAM purified using the purification protocol in the absence of any reducing agent was analysed by SV. The absorbance traces are indicative of the presence of two species, due to two shoulders emerging during sedimentation (Figure 3-3).

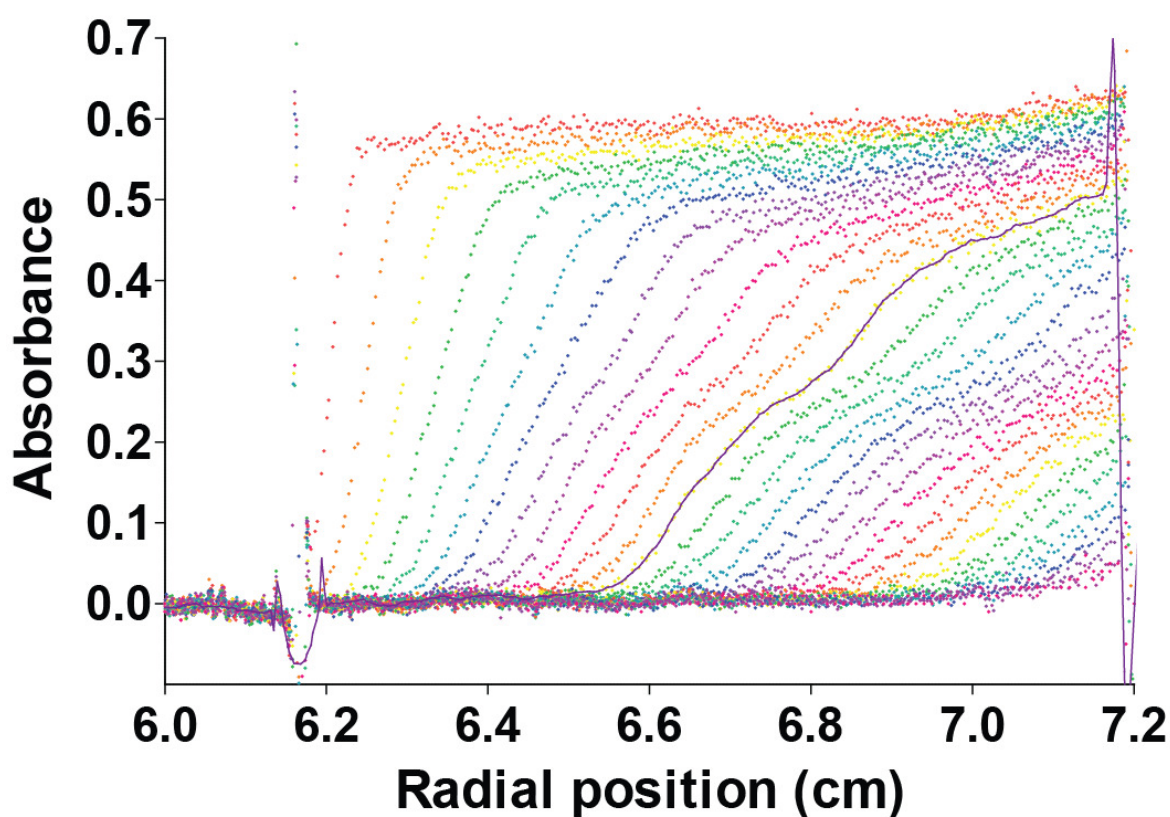


Figure 3-3 Raw absorbance of ECAM during sedimentation velocity experiment.

The absorbance boundary representing the protein at a radial position within the cell moves further out in radial position during sedimentation. The two shoulders present within the purple line indicate multiple species within the solution.

Integrating the particle distribution profiles revealed two distinct populations which were observed at every ECAM concentration and a third species at the highest protein concentrations (Figure 3-4 a). Determination of the sedimentation coefficient at infinite dilution, at twenty degrees in water ($s_{20,w}^0$) for species 1 and species 2 gave values of 7.49 S and 11.20 S respectively (Figure 3-4 b). A third species was noted at the highest concentrations with an $s_{20,w}$ of approximately 15 S. The presence of multiple species in solution confirmed the multiple species previously seen by SEC and SDS-PAGE (Figure 3-1 c, 3-2 b).

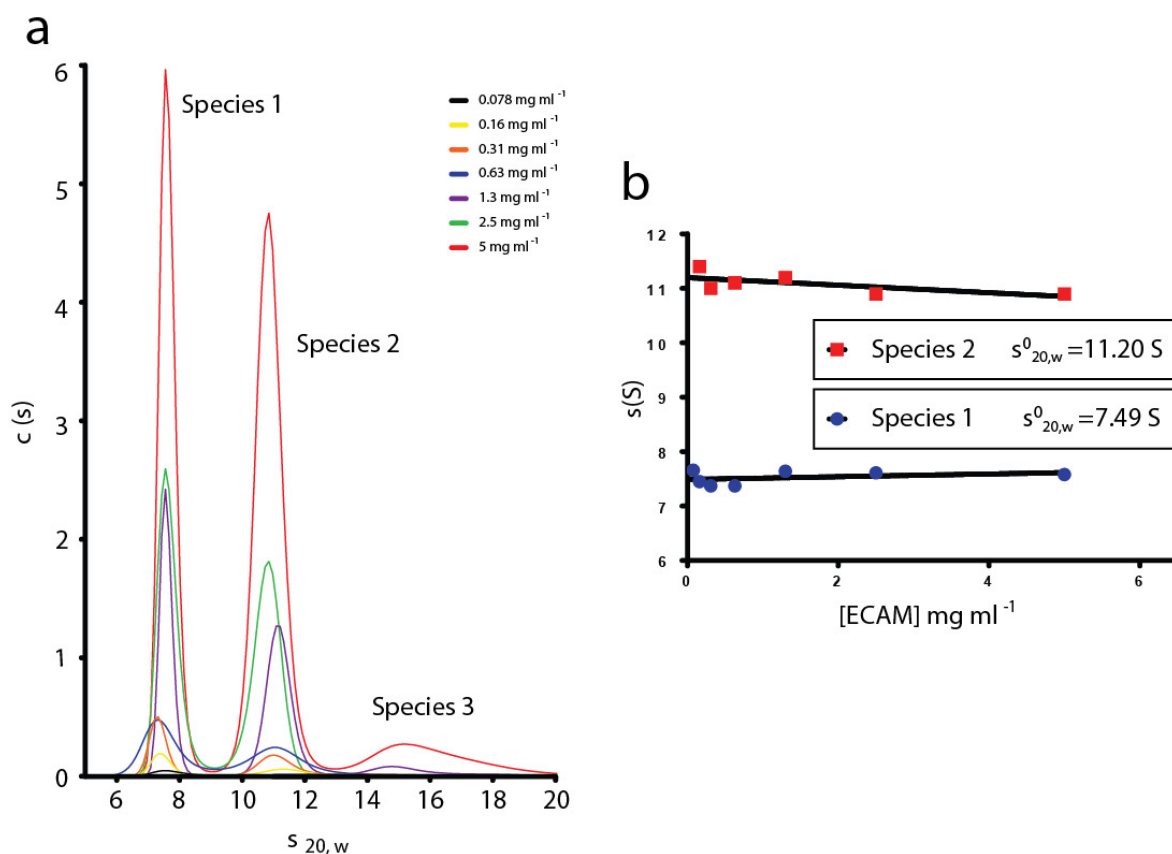


Figure 3-4 AUC analysis of ECAM showing a monomer-dimer mixed population.

(a) continuous $c(s)$ distribution for a range of ECAM concentrations shows three species, possibly monomer dimer and trimer, for which the ratio of % multimer to % monomer increases at higher concentrations. (b) Extrapolation of values to obtain the infinite dilution sedimentation coefficient $s_{20,w}^0$ for ECAM.

When ECAM purified in the presence of TCEP was analysed by SV two species were again present in a similar ratio to those observed for ECAM purified in the absence of TCEP (Figure 3-5 a, 3-6). $s_{20,w}^0$ values for species 1 and species 2 of ECAM purified in the presence of TCEP are 7.53 S and 10.94 S respectively with no third species observed (Figure 3-5 b).

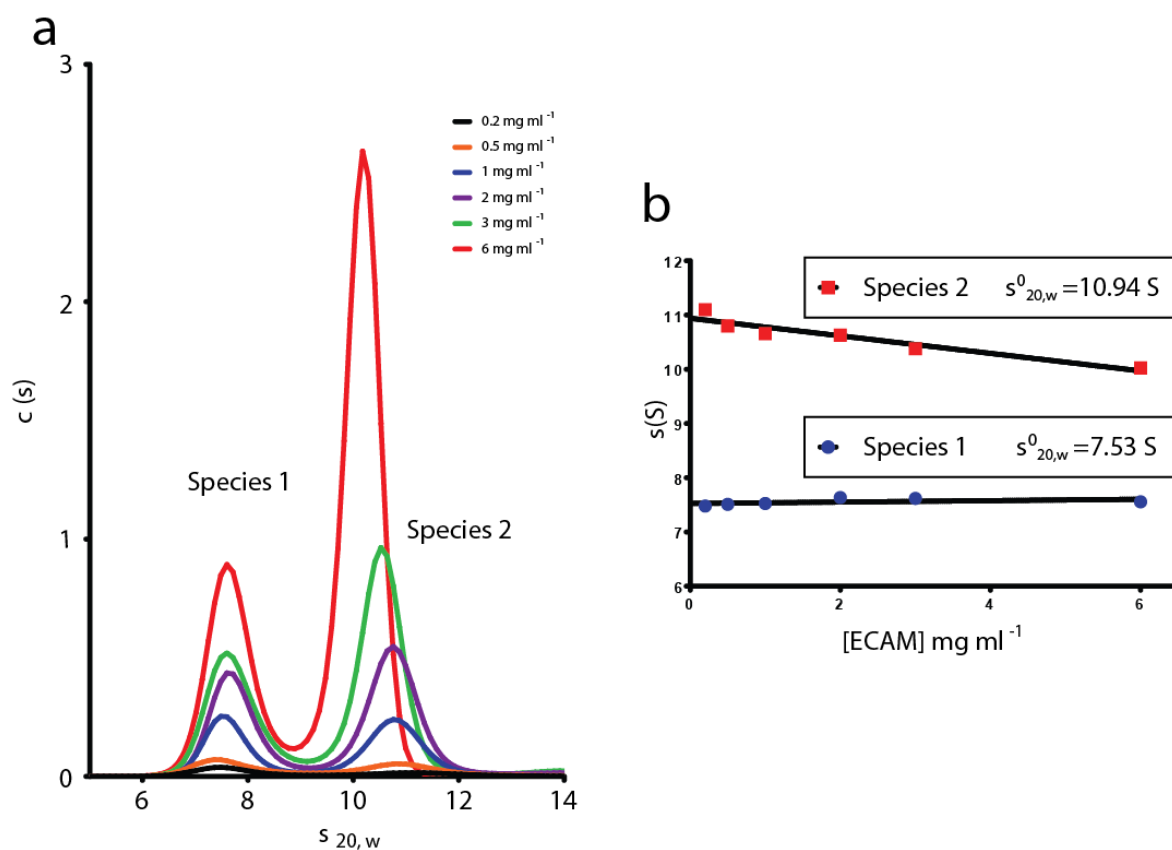


Figure 3-5 AUC analysis of ECAM purified with TCEP showing a monomer-dimer mixed population.

(a) continuous $c(s)$ distribution for a range of ECAM concentrations shows a two species with a gradual decrease in S for species 2 as concentration increases. (b) Extrapolation of values to obtain the infinite dilution sedimentation coefficient $s_{20,w}^0$ for ECAM.

As the concentration of ECAM increased there was an increase in the ratio of species 2 to species 1 with only a minor difference seen with the ECAM purified in the presence of TCEP (Figure 3-6). Using the ratio of dimer to monomer the K_d was calculated as 1.695 μM (using dimer molarity).

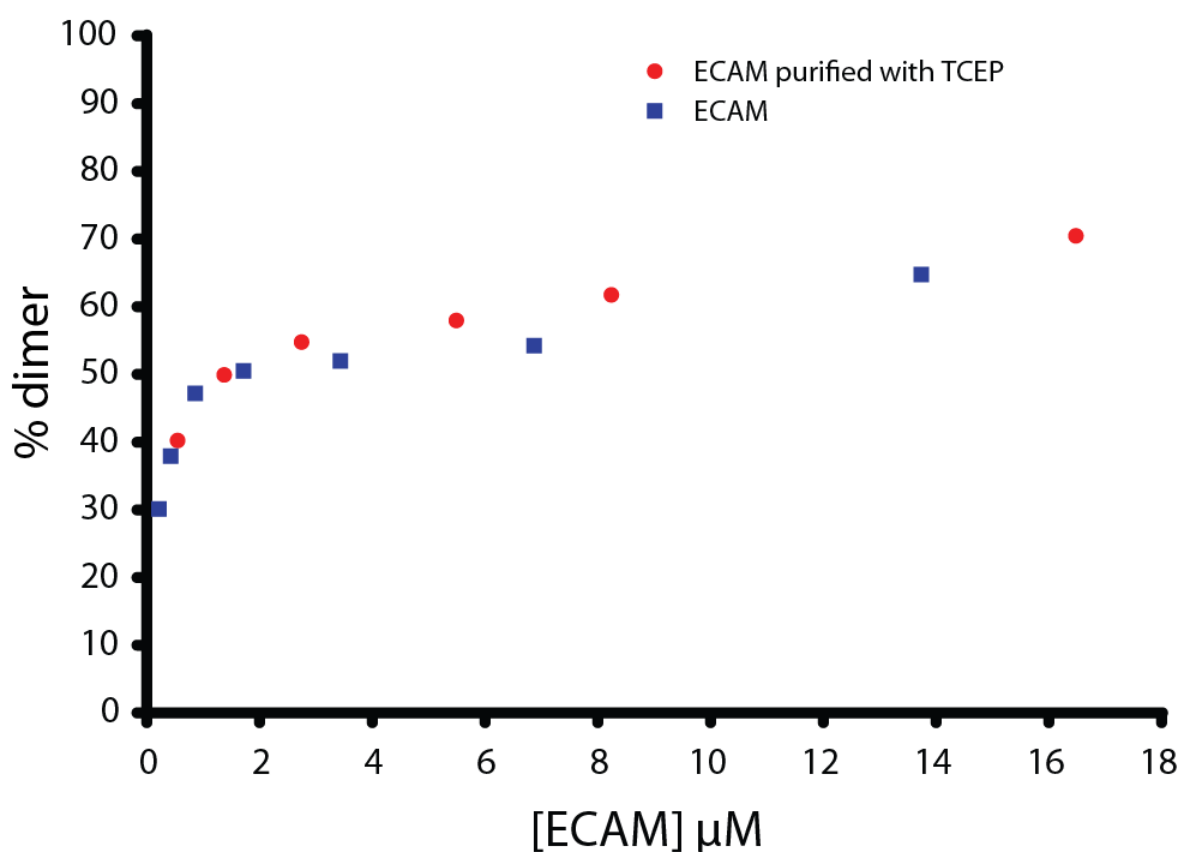


Figure 3-6 Increase in dimer with concentration

Percentage of dimer present in ECAM and ECAM purified with TCEP increases as concentration increases with 65% dimer at 13.7 μM and 70% dimer at 16.5 μM respectively. K_d calculated at 1.695 μM

During the course of this work the crystal structure of an ECAM homologue from *Salmonella enterica* was published¹⁸¹. Although this structure (PDB: 4U48) was submitted to the PDB as a monomer it was noted that there is an entwined crystallographic dimer with a buried surface area of 9155 Å²¹⁴ (Figure 3-7).

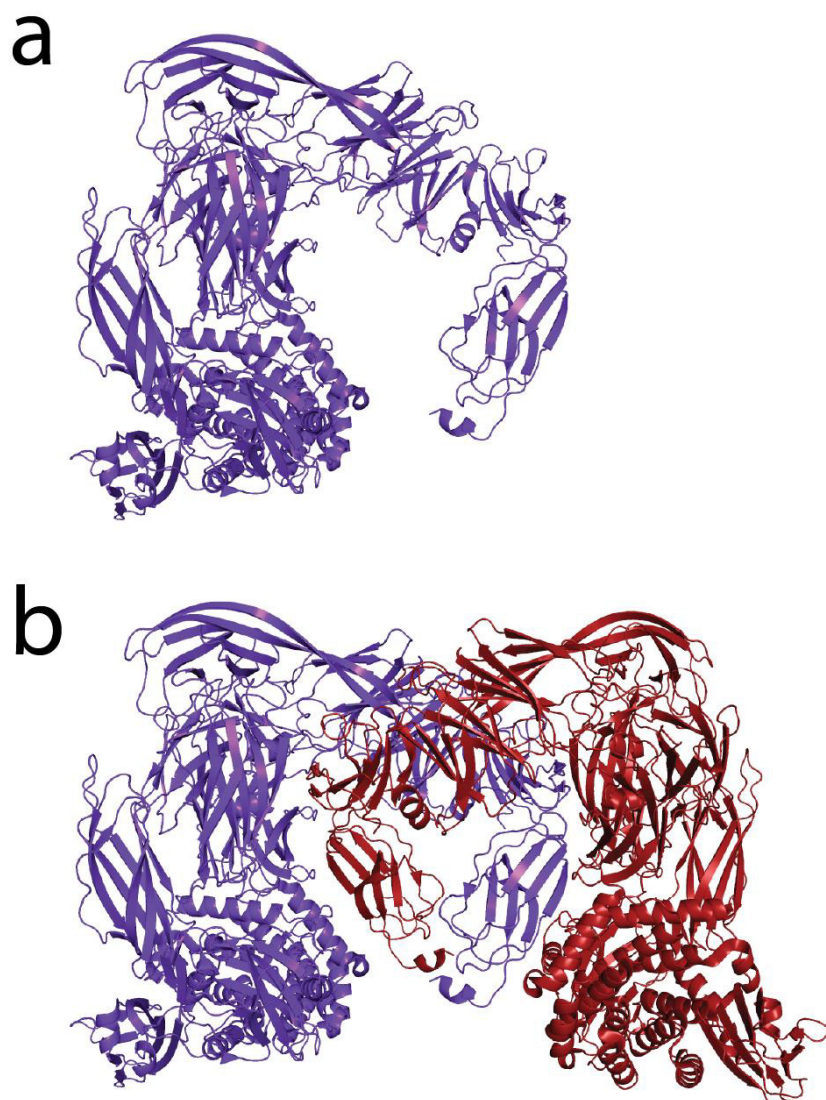


Figure 3-7 Crystal structure of *Salmonella enterica* alpha-2-macroglobulin

(a) monomer form of SaA2M PDB: 4U48 with (b) entwined dimer described as a crystallographic artefact¹⁴.

Using SOMO within Ultrascan, a hydrodynamic bead model of the high resolution monomer and dimer models were generated and used to compute their sedimentation coefficients¹⁶⁶ (Table 3-1). The $s_{20,w}^0$ values determined for monomeric and dimeric SaA2M are 7.99 S and 12.54 S respectively (Table 3-1). The $s_{20,w}^0$ values determined for species 1 of ECAM purified in the presence and absence of TCEP (7.53 S and 7.49 S, respectively), are within 10% of the $s_{20,w}^0$ value calculated for the high resolution model of the monomer SaA2M (7.99 S) (Table 3-1). The $s_{20,w}^0$ values determined for species 2 of ECAM purified in the presence and absence of TCEP (10.94 S and 11.20 S respectively), are within 15% of the $s_{20,w}^0$ value, and 12.54 S as calculated for the high resolution model of the dimer (Table 3-1). The similarity in the sedimentation coefficient values confirms the presence of a monomer and dimer species for ECAM in solution (Table 3-1). However, the lower sedimentation coefficients seen for experimental ECAM data indicate that in solution both the monomer and dimer are less compact and elongated, with a larger surface area resulting in a slower sedimentation compared to the high resolution models. The decrease in sedimentation coefficient for the dimer (species 2) as concentration increases is due to a crowding effect resulting in faster sedimentation (Figure 3-4, 3-5).

Table 3-1 Experimental and computed sedimentation coefficient (S)

		SaA2M crystal PDB:4U48	ECAM	ECAM purified with TCEP
$s_{20,w}^0$	Monomer	7.99 S	7.49 S	7.53 S
	Dimer	12.54 S	11.20 S	10.94 S

3.2.3 Small angle X-ray scattering of ECAM and comparison with high resolution models

Previous SAXS analysis described ECAM as a monomer in solution and modelled an envelope based on this assumption^{68,77}. With the monomer-dimer equilibrium of ECAM noted at even low concentrations, further analysis by SAXS, taking into account the dimeric species could provide further understanding of ECAM composition in solution. When ECAM was analysed by SAXS there was a distinct difference in intensity at low angles when comparing low and high concentration samples that may be due to the monomer-dimer exchange seen by SV (Figure 3-8). Normally small angle, low concentration data can be merged with wide angle high concentration data, however, as SV shows that ECAM is 65% dimer at 6.04 mg ml⁻¹ and 50% at 0.72 mg ml⁻¹ the merging of these data is not valid.

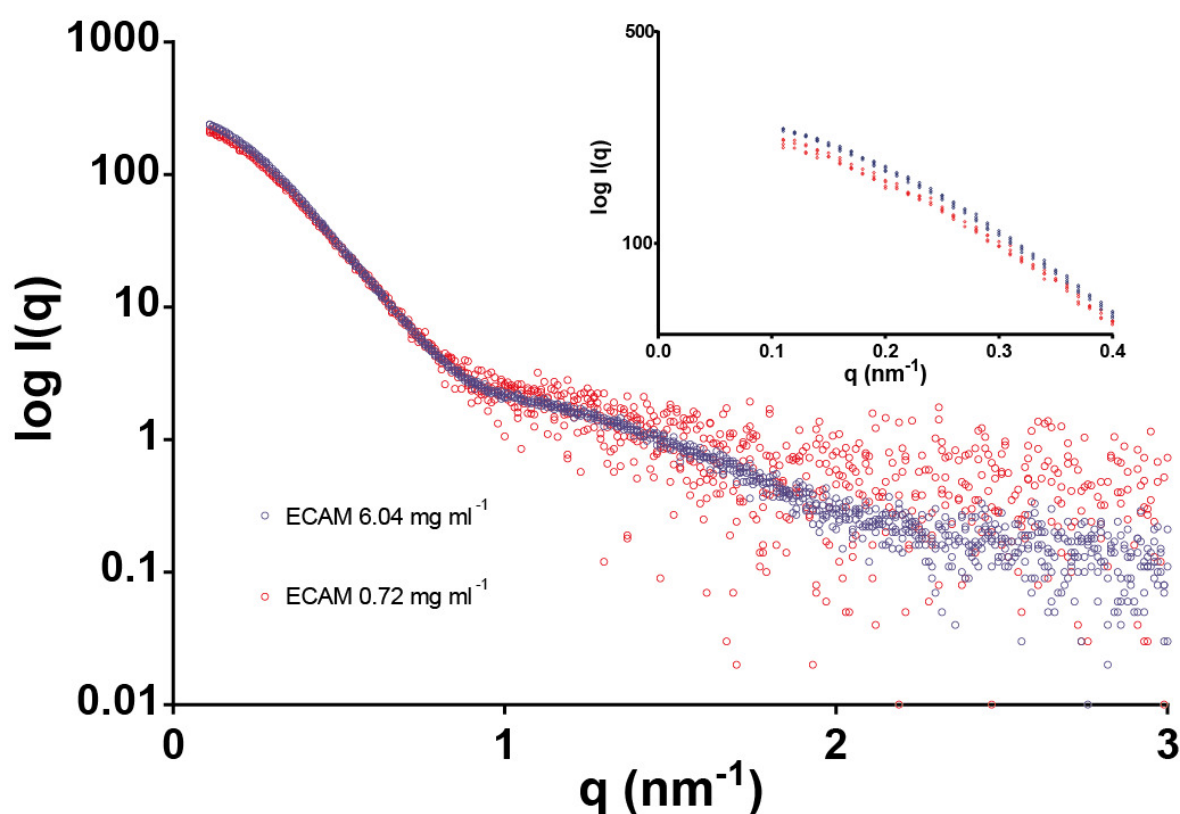


Figure 3-8 Comparison of raw SAXS data for ECAM at high and low concentrations.

The differences seen at small angles are distinct when comparing high and low ECAM concentrations. Inset graph shows a clear difference at low angles.

Guinier analysis of the raw scattering data for ECAM at 6.04 mg ml^{-1} was used to determine the overall size of ECAM calculating the radius of gyration (R_g) as $5.66 \pm 0.009 \text{ nm}$ (Figure 3-9 a, b).

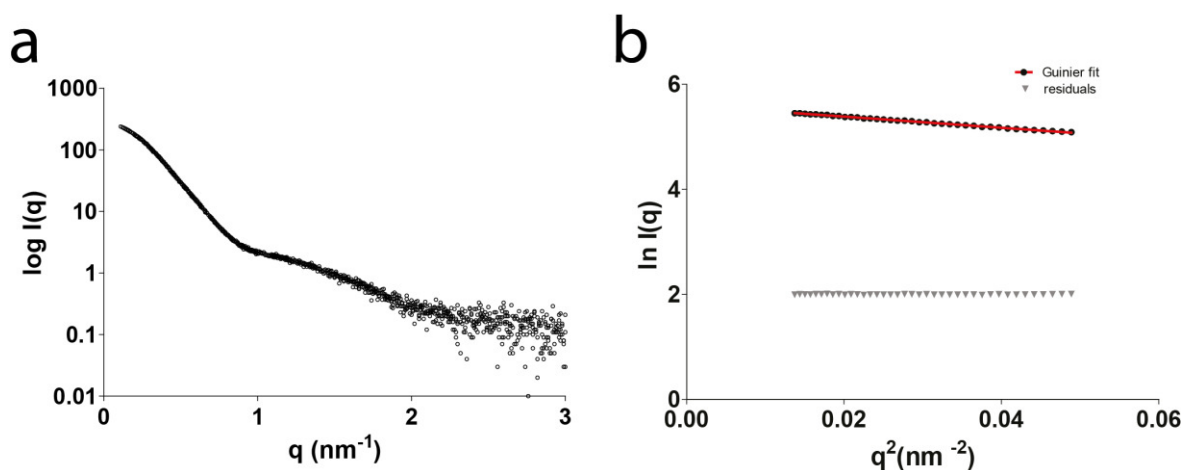


Figure 3-9 Raw SAXS data for ECAM with Guinier fit.

(a) Raw scattering curve for ECAM at 6.04 mg ml^{-1} . (b) Guinier fit to the raw scattering data in the Guinier region determines an R_g of $5.66 \pm 0.01 \text{ nm}$. The linearity of the residuals indicates the absence of aggregation.

The pair-distance distribution function curve, plotting r against $p(r)$, shows distances between electrons within the macromolecule in real space (Figure 3-10 a). From this $p(r)$ plot the most frequent distance observed is 6.0 nm and the maximum distance measured (D_{\max}), taken from where $p(r)$ reaches 0 for a second time, is 18.8 nm . The bell-shaped curve seen for the $p(r)$ plot is normally indicative of a globular protein, however as this is a mixture of two species this is less clear (Figure 3-10 b). The slight hump with a tail remaining low at high q values for the Kratky plot can be interpreted as a multi-domain protein with a rigid structure (Figure 3-10 b).

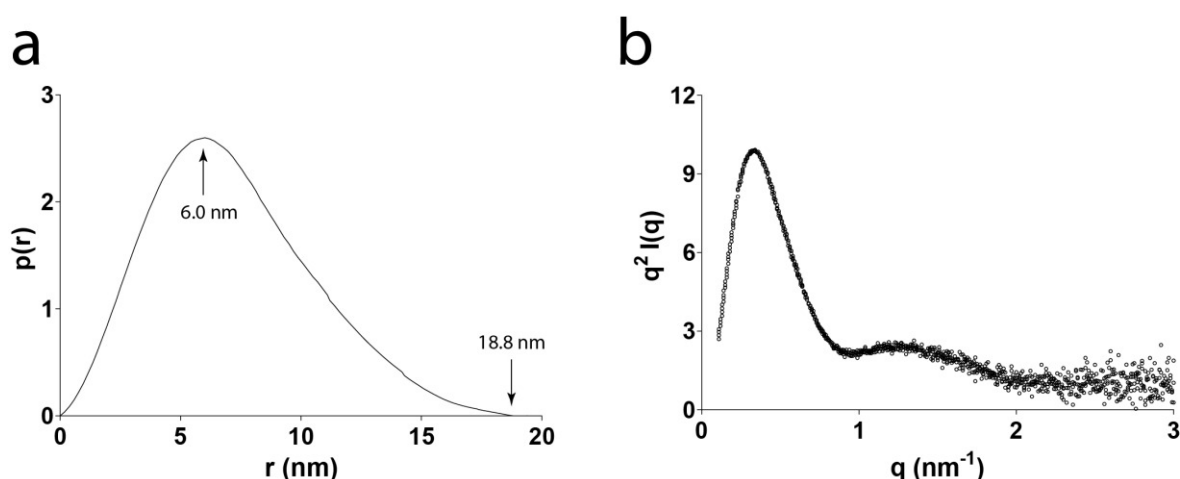


Figure 3-10 Distance distribution function $p(r)$ analysis and Kratky plot for ECAM indicate a globular particle in solution.

(a) $p(r)$ distribution of ECAM at 6.04 mg ml^{-1} has a most frequent distance between electrons of 6.0 nm and a D_{max} of 18.8 nm. (b) The Kratky plot is a bell-shaped curve confirming a globular protein with a hump indicating multiple domains and a tail remaining low suggesting a rigid structure.

The SaA2M monomer and dimer high resolution models previously used to generate hydrodynamic bead models for modelling sedimentation coefficients were likewise used to model SAXS R_g and D_{max} values using SOMO within Ultrascan^{14,166} (PDB:4U48; Figure 3-7, Table 3-2).

Table 3-2 Experimentally determined and computed ECAM biophysical properties

	ECAM (0.72 mg ml^{-1})	ECAM (6.04 mg ml^{-1})	SaA2M monomer PDB:4U48	SaA2M dimer PDB:4U48
MW (kDa)	240 [I_0]	271 [I_0]	181	362
R_g (Guinier) (nm)	5.58 ± 0.042	5.66 ± 0.009	4.24	5.35
D_{max} (nm)	18.5	19.4	13.46	17.14

The zero angle intensity (I_0) can be calculated from the Guinier fit of the SAXS curve and used to approximate the molecular weight of a sample. The molecular weights calculated from I_0 as 240 kDa and 271 kDa for ECAM at 0.72 mg ml⁻¹ and 6.04 mg ml⁻¹ respectively do not match the molecular weight of either monomer (181 kDa) or dimer (362 kDa) (Table 3-2). However, considering that the sample is in equilibrium between monomer and dimer, the average of these monomer-dimer molecular weights being 271.5 kDa is a closer match to that seen experimentally based on I_0 (Table 3-2). The POROD volume can also be used to estimate the molecular mass of a protein by dividing by ~ 1.7 ¹⁷⁶. The calculated POROD volume for ECAM at 6.04 mg ml⁻¹ is 685, when divided by 1.7 gives a molecular mass of 403 kDa just over 10% greater than a dimer, 362 kDa. An observed increase in R_g and molecular weight calculated from I_0 following an increase in concentration can be indicative of aggregation or multimerisation¹⁸². As such the presence of a monomer-dimer equilibrium that increases in dimer composition with concentration follows this trend. The R_g of 5.35 nm and the D_{max} of 17.14 nm for the SaA2M dimer high resolution model is not a good fit to the experimentally determined R_g of 5.66 nm and D_{max} of 19.4 nm for ECAM (Table 3-2). The larger R_g and D_{max} seen experimentally indicate a more elongated molecule than the high resolution model would suggest.

CRY SOL was used to calculate SAXS curves for the SaA2M monomer and dimer high-resolution models (Figure 3-11)¹⁸⁰. The CRY SOL fit of the monomer high-resolution model to the raw scattering curve for ECAM at 6.04 mg ml⁻¹ shows that there are distinct differences in the small angle and mid-range angle with a χ of 11.867 (Figure 3-11)¹⁸⁰. The dimer high-resolution model CRY SOL fit has a closer match to the raw scattering curve than the monomer model, with a distinct match in the small and the wide angle regions and a χ of 5.471 (Figure 3-11)¹⁸⁰. The dissimilarity between the CRY SOL curves for both monomer and dimer high-resolution models and raw scattering curve are expected as the protein solution is a mixture of both monomer and dimer (Figure 3-11).

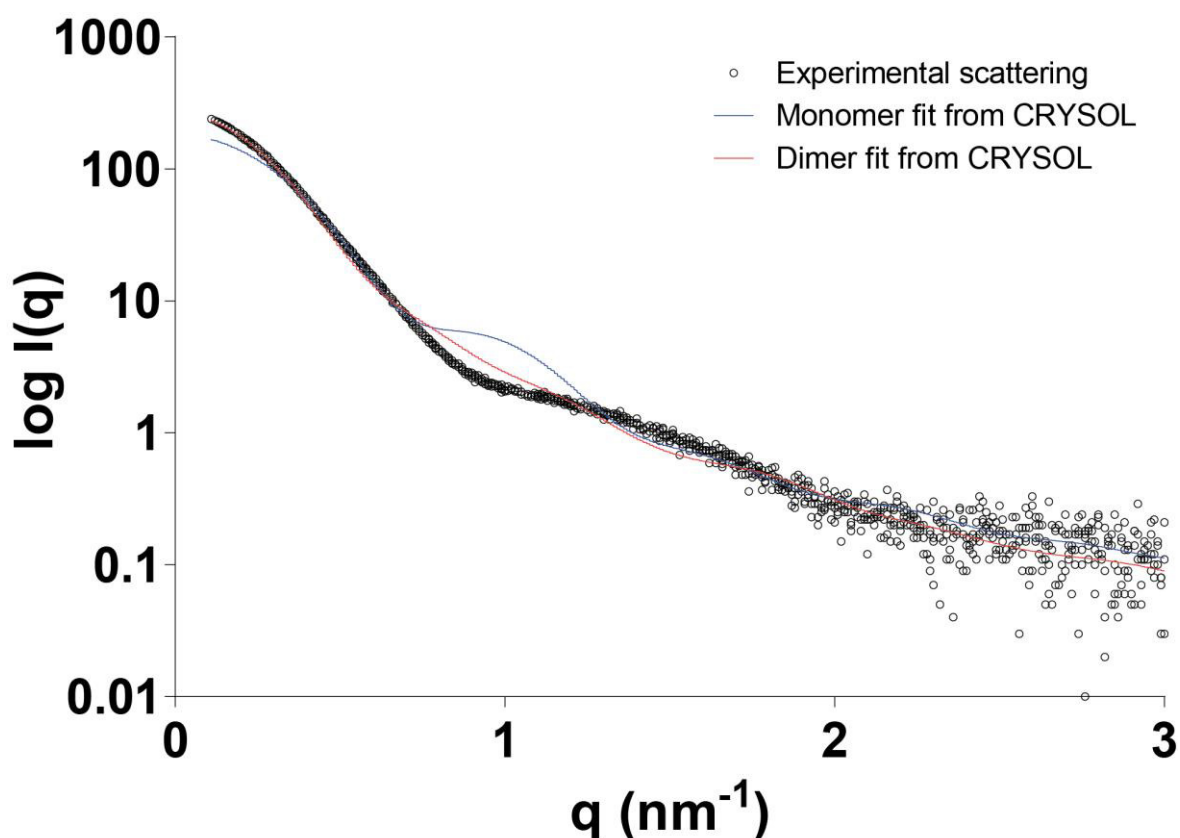


Figure 3-11 Comparisons between ECAM experimental scattering and CRYSOLE generated unreacted SaA2M monomer and dimer high resolution model curves.

ECAM experimental scattering (black circles, 6.04 mg ml^{-1}) compared with monomer (blue) and dimer (red) CRYSOLE curves of SaA2M high resolution model PDB: 4U48^{14,180}.

Ab initio model generation of particles in solution based on the scattering of X-rays can be performed using various programs as part of the ATSAS suite of programs resulting in an envelope representing the low resolution model of the protein¹⁷⁶. However, generation of envelopes was not deemed suitable as the sample is a mixed species of monomer-dimer.

3.3 Discussion

Previous work has suggested that ECAM forms multimers on activation by proteases but is monomeric in its unactivated form⁶⁷. In this work, when ECAM was analysed by SEC there was clearly two species present when purified in the absence of any reducing agent (Figure 3-1). This monomer and dimer population is also seen by SDS-PAGE (Figure 3-1). SV analysis confirms the multiple species seen with sedimentation coefficients are similar to the simulated sedimentation coefficients from high resolution models (Table 3-1). SV analysis also shows that as the concentration of ECAM increases, more dimer is present (Figure 3-6). Upon SAXS analysis the increase in R_g and MW calculated from I_0 also indicate an increase in multimerisation as concentration is increased (Figure 3-8, Table 3-2). When sedimentation coefficients are computed for the monomer and dimer high resolution models for SaA2M, the higher values than experimental species indicate that ECAM is more elongated in solution (Figure 3-12, Table 3-2).

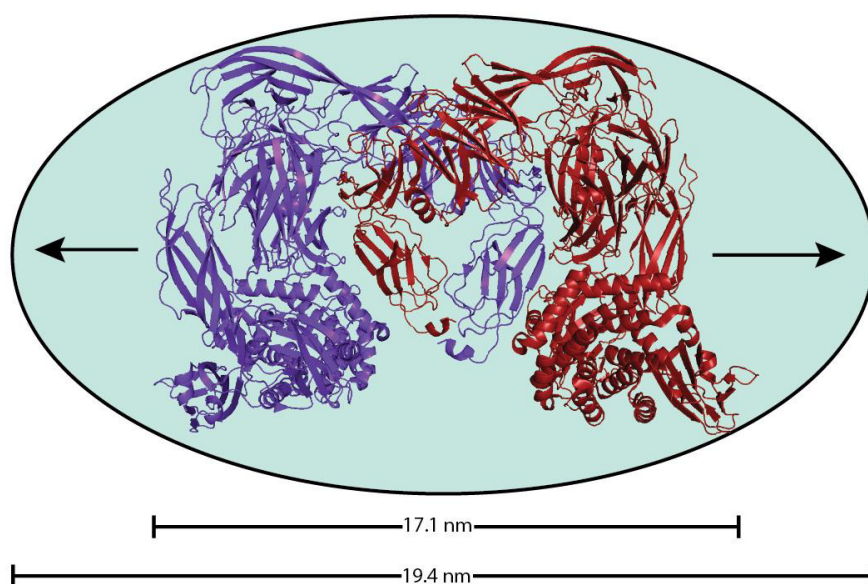


Figure 3-12 ECAM dimer has a larger conformation than SaA2M high resolution model

The lower sedimentation coefficient and higher R_g and D_{max} for experimental ECAM data values indicate that ECAM dimer has a larger conformation than seen in the crystallographic dimer for SaA2M.

Further, the higher R_g and D_{max} seen experimentally by SAXS when compared to SaA2M model computed R_g and D_{max} values again indicates a less compact elongated formation (Figure 3-12, Table 3-2). The characterisation of ECAM in solution has shown that the native protein that has not been reacted with protease is in a monomer-dimer equilibrium that increases in dimer ratio with increase in concentration.

This monomer-dimer mixed species can be observed by SEC, SDS PAGE, AUC and confirmed through CRY SOL comparison of a 'crystallographic' dimer with SAXS raw scattering curve. Further analysis of the high resolution models through simulated hydrodynamic parameters confirms the presence of monomer-dimer mixed species seen through AUC and SAXS¹⁸⁰. The differences seen in CRY SOL curve of SaA2M dimer and experimental scattering curve and envelope generated can be explained by the monomer-dimer equilibrium shown to be present and the flexibility within a dimer and the fluctuation that would be expected between monomer and dimer forms (Figure 3-11)¹⁸⁰. The dimer form that is seen at high concentrations of protein may form under *In vivo* conditions due to crowding by heterogeneous proteins. ECAM is also a membrane anchored protein and when movement is limited to a two dimensional plane rather than three dimensions seen in solution it would be expected that dimerization could occur at lower protein concentrations than observed in solution (Figure 3-13).

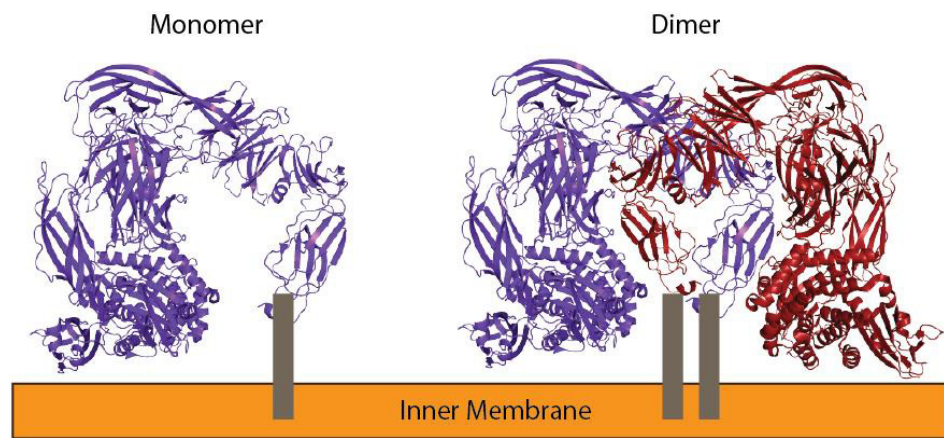


Figure 3-13 ECAM model of interaction with inner membrane within periplasm.

4 Crystallisation and structure model building of protease cleaved ECAM

4.1 Introduction

Although, a structure for methylamine activated HA2M has been solved, this is a low resolution structure and provides little information on the mechanism of activation of α 2Ms. In order to understand this mechanism we attempted to elucidate the structural changes that occur on protease activation of ECAM. As elastase had been used previously for characterisation of ECAM and has known cleavage targets within the bait region of ECAM this was considered a good candidate for activating ECAM.

4.2 Results

4.2.1 Crystallisation of porcine elastase activated ECAM

Purified ECAM was reacted at a 1:1 molar ratio with porcine elastase and purified by SEC with fractions from the two largest peaks being pooled separately for crystallisation trials (Figure 4-1). The fractions from the two main peaks are likely to be digestion products of ECAM as they elute after full length ECAM. The fractions from these two peaks were individually concentrated to 16 mg ml⁻¹ and crystallisation trials were set up using the Molecular Dimensions crystallisation screens MORPHEUS®, MIDAS™, JCSG-plus, and PACT premier. Crystal growth was only observed for protein from peak 1, in one condition from the MIDAS™ screen with a reservoir solution containing 0.1 M potassium chloride, 0.1 M HEPES, 25% SOKALAN® CP 7, pH 7.0. For optimisation of crystallisation, the concentrations of KCl, HEPES, and SOKALAN® were varied as was the pH.

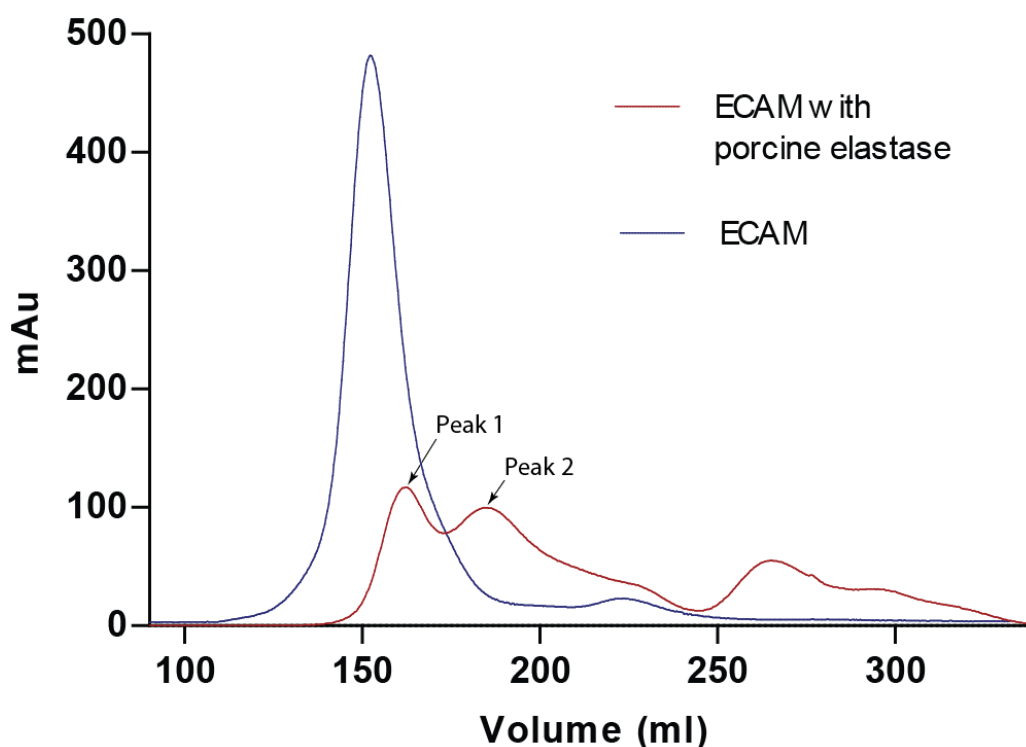


Figure 4-1 Size exclusion chromatography of ECAM and ECAM reacted with porcine elastase.

ECAM purified by size exclusion chromatography shown in blue and ECAM reacted with elastase shown in red. Numbered peaks highlight the fractions collected from ECAM reacted with elastase and subsequently used for crystallisation trays. ECAM used had previously been purified by Ni^{2+} affinity chromatography in the presence of TCEP.

The condition resulting in the crystal growth with the largest crystals was similar to the initial hit with only a change in pH to 7.5 (Figure 4-2 a). Variation of ratio of protein to reservoir solution was also trialled for both SEC protein peaks although crystals were only seen in the initial screen for peak 1. A 1:1 ratio of protein solution and reservoir solution was found to be the best ratio for crystal growth with crystal appearance within 2 days for protein from peak 1. Crystals with similar morphology appeared after 3 weeks for protein from peak 2. Upon the appearance of crystals, growth significantly slowed after a week, after which time crystals were tested for diffraction on the home beam.

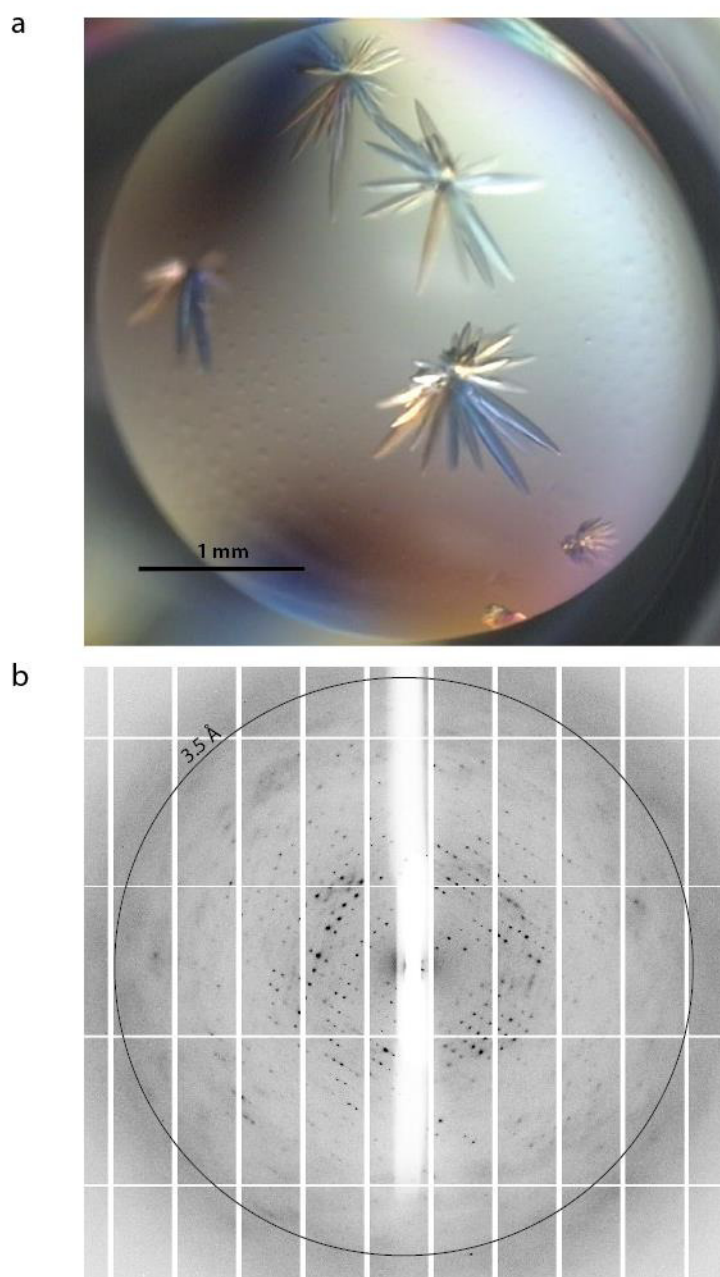


Figure 4-2 Crystals from ECAM reacted with porcine elastase and their diffraction.

(a) Individual rods were separated from the clusters of crystals grown and used for data collection. (b) X-ray diffraction image collected at DIAMOND synchrotron light source from crystals of unlabelled porcine elastase reacted ECAM; the data set obtained extended to 3.5 Å.

As the intention of reacting ECAM with proteases before crystallisation was to obtain crystals of ECAM bound to protease, crystals were picked and run on SDS-PAGE and bands analysed by trypsin digest and liquid chromatography (LC) followed by mass spectrometry (MSMS) of peptides (Figure 4-3). SDS-PAGE showed a faint band at 120 kDa and three stronger bands at 90, 75 and 27 kDa. LCMSMS of the peptides obtained by trypsin digestion of the bands indicated that there was no porcine elastase present. The band at 120 kDa showed peptide hits for the whole protein minus MG domains 1, 2, 3, and 7. The bands at 90 kDa and 75 kDa showed peptide hits for all domains between MG8 to the CTMG domain. The band at 27 kDa had peptide hits for domains MG4 and MG5 (Figure 4-3).

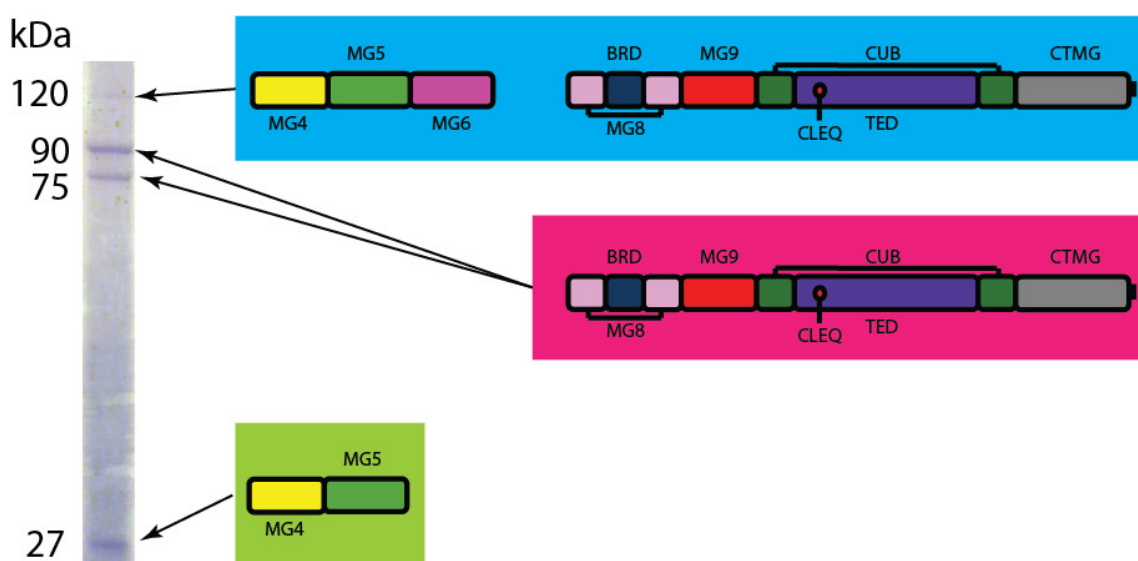


Figure 4-3 Elastase treated ECAM crystals on SDS-PAGE and LC MSMS peptide coverage from bands.

Peptide coverage was found from the corresponding domains upon trypsin digestion and LC MSMS. The band at 120 kDa (blue) had peptide coverage from the all domains seen in the crystal structure and the bands at 90 and 75 kDa (pink) had peptide coverage from the N-terminus seen in the crystal structure. The band seen at 27 kDa (green) had peptide coverage from domains MG4 and 5.

The standard reservoir solution that produced crystal growth provided no cryoprotection and so it was mixed with various cryoprotectants. Glycerol, ethylene glycol, and PEG (various small molecule sizes) were trialled at various ratios, however all of these cryoprotectants resulted in the solution obtaining an opaque viscous appearance. As this may have a detrimental effect on cryoprotection or stability of crystals, sugars were then tested as cryoprotectants. The reservoir solution was mixed with either sucrose or xylitol resulting in a saturated sugar reservoir solution. The addition of xylitol was successful at providing cryoprotection with a 3:2 ratio of xylitol saturated reservoir solution to reservoir solution. Crystals were picked and briefly dipped in the xylitol cryoprotectant solution before being tested for diffraction on the home source with the best diffraction of $\sim 8 \text{ \AA}$ found using larger crystals with $>500 \text{ \mu m}$ in length and $>60 \text{ \mu m}$ in thickness. It was found that larger crystals cracked upon being added to the cryoprotectant and this could be avoided by soaking crystals in solutions that gradually increases xylitol content until the 3:2 ratio that provided cryoprotection. From the diffraction data obtained on the home beam it was possible to determine a space group of H3 and cell dimensions of 176.06, 176.06, 161.13 ($a, b, c \text{ (\AA)}$), 90, 90, 120 ($\alpha, \beta, \gamma \text{ (}^\circ\text{)}$). Multiple crystals were stored for future data collection at the DIAMOND synchrotron light source (Figure 4-2 b). Data were collected for ECAM crystals at the I02, I03 and I24 beamlines at Diamond Light Source (Didcot, United Kingdom) at 100 K at the selenium K-edge ($\lambda = 0.97939 \text{ \AA}$) using a Pilatus 6M detector.

4.2.2 Phasing and model building of porcine elastase activated ECAM

Upon successfully collecting a complete data set to 3.5 Å resolution for native porcine elastase activated ECAM crystals, molecular replacement phasing of this data was attempted using previously published crystal structures of α 2M. Although only having 12% sequence identity between ECAM and α 2M, we decided that as the expected activated domain orientation should be similar this may work for phasing our data through molecular replacement, however this was unsuccessful. The Phyre 2 protein fold recognition server was used to generate a model that had the activated domain structure of α 2M with 100% sequence identity to ECAM, however, this was also unsuccessful¹⁸³. For phasing heavy metal derivatisation was also attempted to allow single-wavelength anomalous diffraction (SAD) or multiple-wavelength anomalous diffraction (MAD). Crystals were soaked in multiple gold, platinum and palladium salts and although crystals changed in colour upon soaking without obvious damage to the crystals, no diffraction could be detected. Using selenium labelled protein was the next method attempted to provide phase information.

A methionine auxotrophic strain of *E. coli* (T7 Express Crystal Competent *E. coli* (New England BioLabs)) supplemented with selenomethionine and other essential amino acids were used to express ECAM. The previous method of crystallisation after reacting ECAM with porcine elastase and purifying by SEC was attempted, however this was unsuccessful. After several purifications still resulting in no crystals, unreacted selenomethionine labelled ECAM (SeMetECAM) was mixed at a ratio of 100:1 ratio of SeMetECAM to porcine elastase using the MORPHEUS®, MIDAS™, JCSG-plus, and PACT premier Molecular Dimensions crystallisation screens used previously. Crystals were formed with a similar morphology as seen previously under identical crystallisation conditions (0.1 M KCl, 0.1 M HEPES, 25% SOKALAN® CP 7, pH 7.0) as for porcine elastase reacted unlabelled ECAM protein. This *in situ* proteolysis was repeated with the optimised condition (0.1 M KCl, 0.1 M HEPES, 25% SOKALAN® CP 7, pH 7.5) with successful crystal growth. These crystals grew as pointed rods as large as 1 mm long with 0.1 mm diameter and diffracted to 3.65 Å (Table 4-1). Several high redundancy SAD data

sets were collected from one well diffracting crystal and processed, scaled and merged using programs *XDS* and *AIMLESS* from the *CCP4* suite of programs^{167,168,170}.

Table 4-1 Data collection and refinement statistics (single-wavelength anomalous diffraction) for protease cleaved ECAM

PDB: 4RTD	
Data Collection ^a	
Space Group	<i>H3</i>
Cell Dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	176.06, 176.06, 161.13
α , β , γ (°)	90, 90, 120
Resolution (Å)	46.87-3.65 (4.0-3.65)
Solvent Content (%)	69
No. of unique observations	20753 (4991)
CC _{1/2}	0.998 (0.675)
<i>R</i> _{merge} (%)	39 (378.0)
<i>R</i> _{p.i.m} (%) ^b	7.7 (74.4)
Mean <i>I</i> / σ (<i>I</i>)	13.2 (2.1)
Completeness (%)	99.9 (99.9)
Redundancy	26.6 (26.7)
Anomalous completeness (%)	99.9 (99.3)
Anomalous redundancy	12.8 (12.8)
DelAnom correlation between half-sets	0.335 (0.011)
Mid-slope Anomalous Normal Probability	1.245
Refinement statistics	
<i>R</i> _{work} / <i>R</i> _{free} (%)	17.7/23.8
No. of atoms	8699
RMSD of bond lengths (Å)	0.01
RMSD of bond angles (°)	1.53
Mean/Wilson plot B-value (Å ²)	142.1/100.7
Ramachandran plot (%) ^c	
<i>Favoured/Allowed/Outliers</i>	90.7/7.9/1.3

Data collected from one crystal

^a Values in parentheses refer to the highest resolution shell.

^b $R_{p.i.m} = \sum_{hkl} [1/(N - 1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$

^c Percentages of residues in favoured/allowed regions calculated by the program *Molprobit*¹⁷⁵

Upon processing the data and building the model, statistics for data collection, experimental phasing and refinement are presented in Table 4-1. The final electron density map is of very high quality for a resolution of 3.65 Å, with sidechains clearly defined (Figure 4-4). This high quality is due to the phasing obtained from a strong highly redundant data set (Table 4-1).

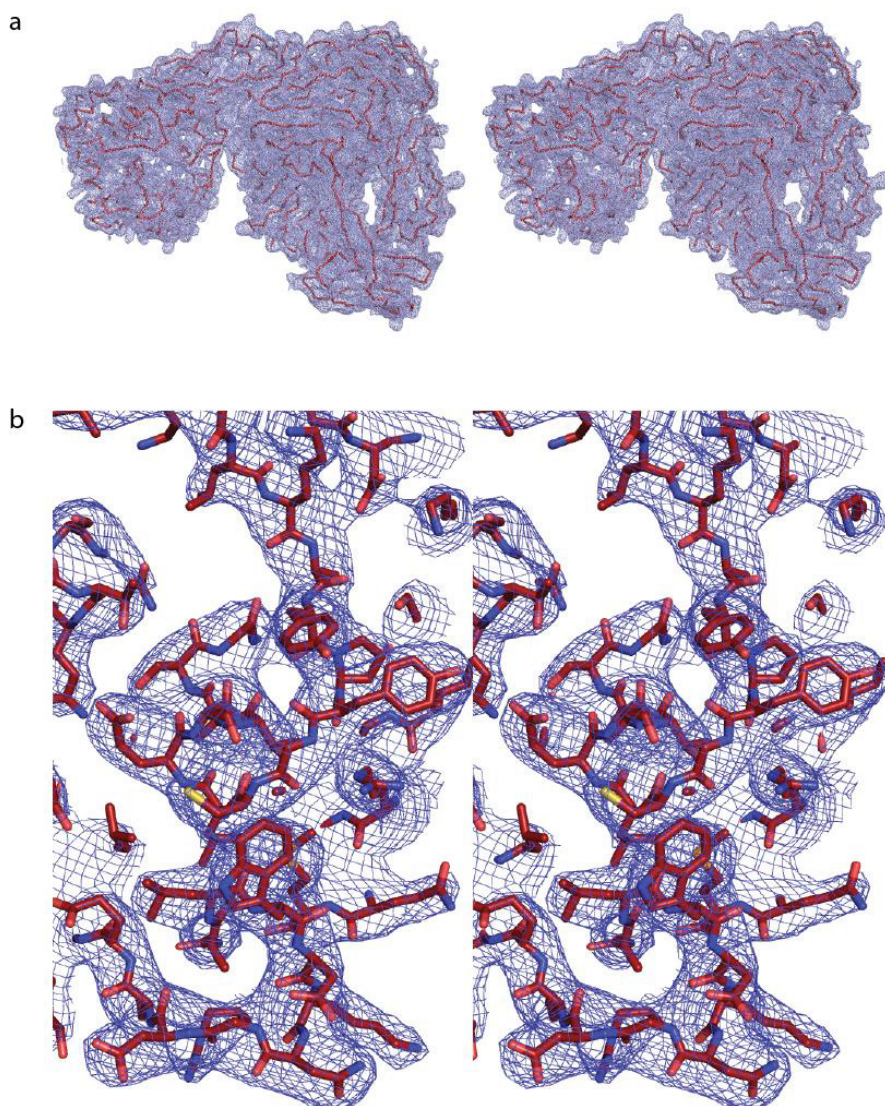


Figure 4-4 Stereo image structure within 1 σ electron density envelope.

(a) Electron density matches the backbone of whole protein structure within electron density envelope drawn at 1 σ level. (b) Thioester region within electron density envelope drawn at 1 σ level with clear sidechains apparent at 3.65 Å.

Within our protease-activated ECAM structure the distance between the sulphur atom of C1187 and the γ -carbon of Q1190 is 4.6 Å indicating that the thioester bond may not be intact (Figure 4-5). However, we cannot rule out the possibility that there is a mixed population of molecules some of which possess an intact thioester bond. Due to this ambiguity we have represented the thioester without a covalent bond between C1187 and the γ -carbon of Q1190 and have also omitted the oxygen from the deaminated glutamine that would be formed on hydrolysis of the thioester bond (Figure 4-5).

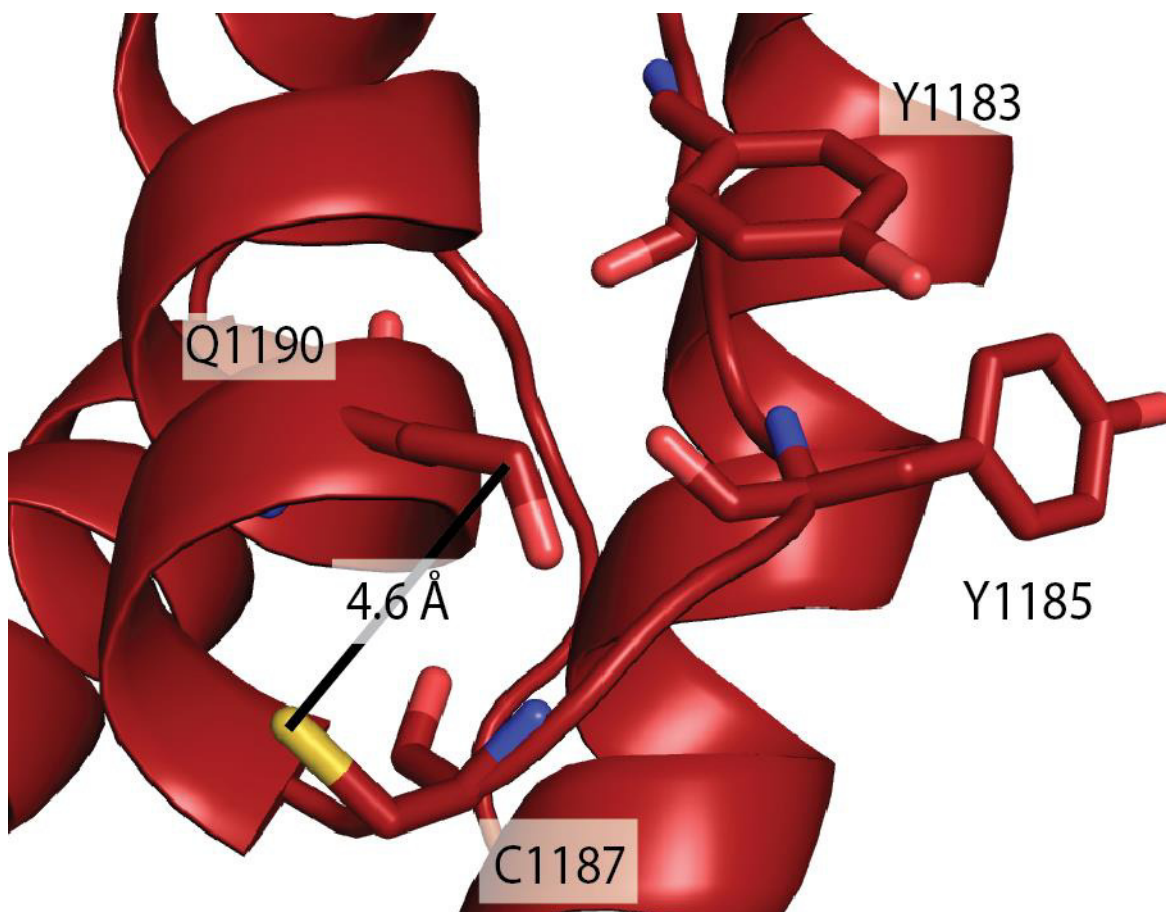


Figure 4-5 Cleaved thioester of protease-cleaved ECAM.

The S-C γ distance between C1187 and Q1190 is seen to be 4.6 Å, suggestive of a cleaved thioester bond.

4.2.3 Overall structure of protease-activated ECAM

Upon completion and validation of the model for protease-activated ECAM, the number of Ramachandran outliers (1.3%) remaining was appropriate for a crystal structure with a resolution of 3.65 Å. Although R_{merge} and R_{pim} values were high, this can be explained by the highly redundant dataset used, but with a $CC_{1/2}$ of 0.675 in the highest shell these data used were judged to be acceptable^{184,185}. Similar to the domain architecture of native SaA2M, un-cleaved ECAM consists of 10 MG domains, with a bait region that is predicted to be largely disordered found within MG8, a TED that houses the reactive thioester bond and a CUB domain (Figure 4-6 a, b). Elastase-cleaved ECAM, (Figure 4-6 b) adopts a conformation similar to that of methylamine-activated human α 2M (Figure 4-7 a), but distinct from the unactivated form of SaA2M (Figure 4-7 b). Despite the low sequence identity between human α 2M and ECAM (12%), the r.m.s.d between α -carbons (when using equivalent domains within ECAM structure) for these proteins is 14.1 Å while for SaA2M and ECAM, which share 82% sequence identity, is 22.1 Å. Most notably, in the structure of the protease cleaved ECAM, interactions between the TED and CTMG domain, which protects the thioester bond in unactivated SaA2M, are not present. Instead, the thioester region of TED is solvent exposed and faces the expected location of the attacking protease, as it would be positioned when cleaving the bait region of ECAM (Figure 4-6 b). Electron density for elastase, in addition to that for MG domains 1, 2, 3 and 7, was absent in $2Fo-Fc$ maps of cleaved ECAM as was electron density for 20 amino acid residues (R923 – L942) within the bait region (Figure 4-6). Although there is space within the crystal packing to accommodate the presence of elastase the lack of any elastase observed upon SDS-PAGE of crystals confirmed the lack of elastase presence (Figure 4-3). Porcine elastase I has been shown to have specificity of cleavage between alanine-alanine and alanine-glycine residues with porcine elastase II having specificity between leucine-alanine, leucine-glycine, phenylalanine-alanine, phenylalanine-glycine, tyrosine-alanine and tyrosine-glycine residues¹⁸⁶. The absence of domains MG1,2,3 and MG7 as well as elastase from the crystal structure could be due to elastase not being inhibited by ECAM and proceeding to cleave in various places beyond the bait region. The bait region has a RLAALR motif between R941 and R946 within the

bait region with a likely cleavage between L942 and A943 as no density is present for residues prior to A943. This bait region cleavage site is very close to the cleavage site seen in the trypsin cleaved ECAM structure (PDB: 4ZIQ) between residues G948 and G949¹⁸⁷. There is a likely cleavage location in the linker of MG3 and MG4 between residues A379 and G380 with electron density appearing shortly after at S385. With the absence of MG7 likely cleavage locations for porcine elastase could be between G751 and A752 and L852 and G853.

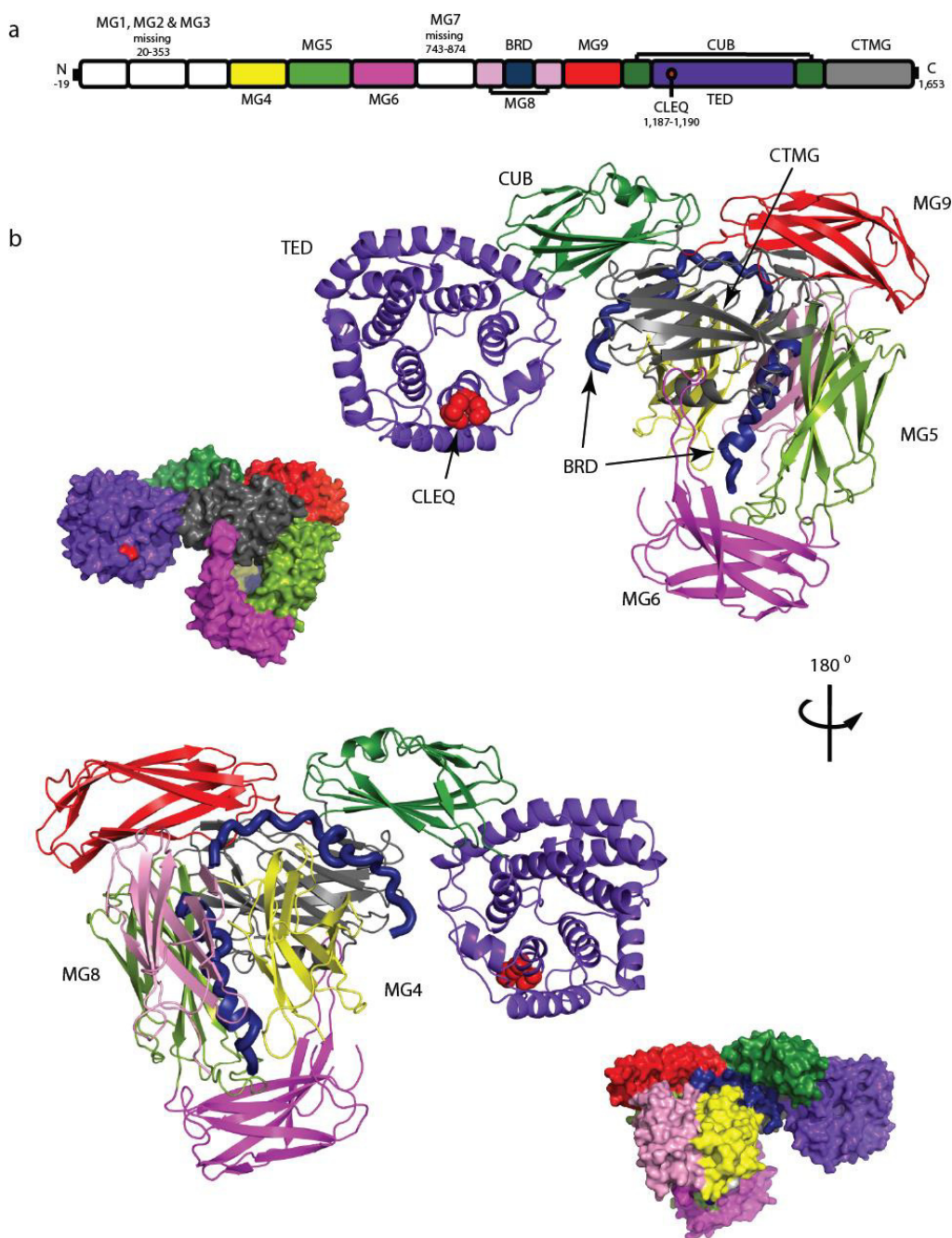


Figure 4-6 Crystal structure of porcine elastase cleaved *Escherichia coli* $\alpha 2M$ (ECAM).

(a) Schematic representation of the 13 domains of ECAM showing the macroglobulin domains (MG) including the C-terminal MG (CTMG) domain, bait region domain (BRD), complement protein subcomponent domain (CUB), and the thioester domain (TED) containing the CLEQ motif. White boxes represent domains in full length ECAM however not present within crystal structure. (b) The structure of elastase cleaved ECAM with the individual domains coloured as in (a) shown in two views from opposing directions. Smaller van der Waals surfaces in both views are also presented. In the lower view the alpha-helical TED is orientated with the CLEQ thioester (drawn as red van der Waals spheres) positioned above the pocket, which is thought to accommodate the attacking protease.

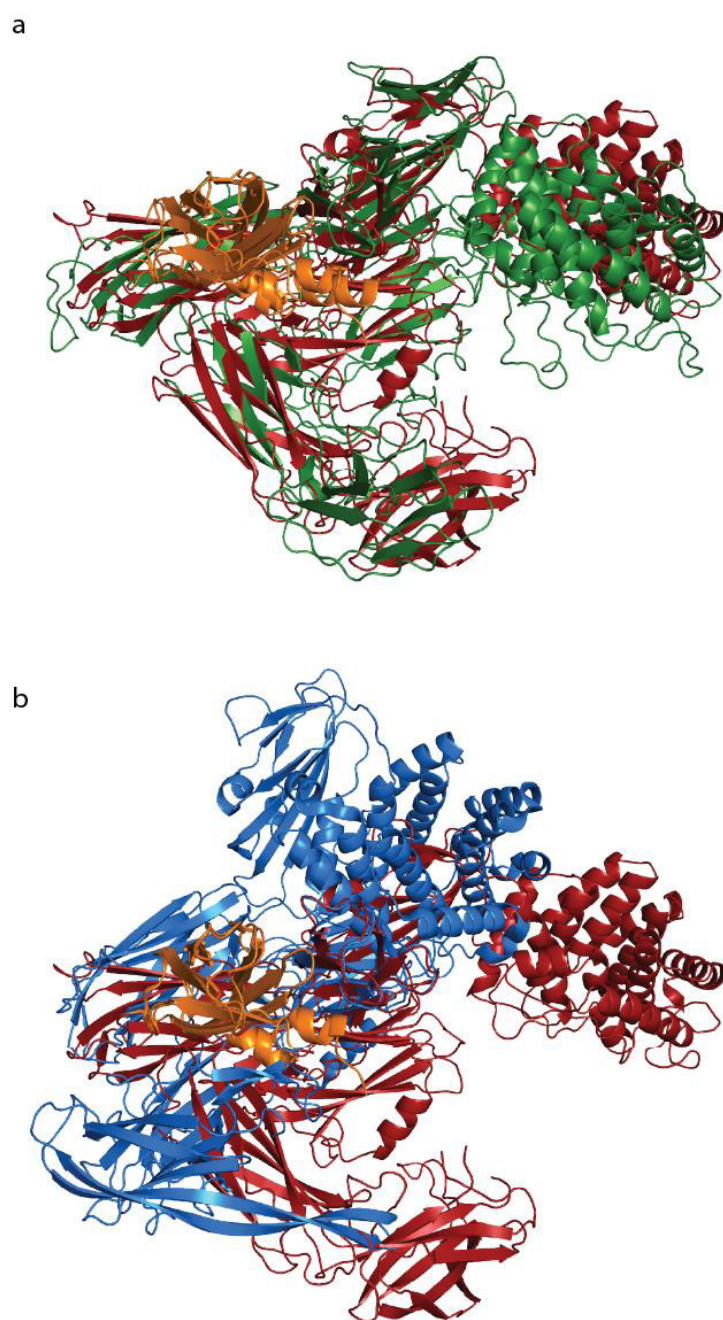


Figure 4-7 Comparison between trimmed ECAM, SaA2M, and HA2M.

(a) Structural alignment of methylamine treated human α 2M (PDB: 4ACQ monomer trimmed to domains present in cleaved ECAM) and elastase cleaved ECAM (PDB: 4RTD) in green and red, respectively. (b) Structural alignment of native SaA2M (PDB: 4U48 trimmed to domains present in cleaved ECAM) and elastase cleaved ECAM (PDB: 4RTD) in blue and red, respectively. Structural alignments were performed using the MG domain containing the bait region shown in orange.

4.2.4 Interaction of the BRD with CTMG

Although it is known that protease cleavage of the largely disordered bait region of both human and bacterial α 2Ms gives rise to a large conformational rearrangement and activation of the thioester bond, the mechanism through which this is mediated was unclear^{20,77}. However, the structure of protease cleaved ECAM provides a putative mechanism for this process. Thioester bond release may be achieved by the untethering of a large region of an unstructured polypeptide chain upon cleavage of the BRD that forms new interactions with the CTMG domain, thereby preventing the protective interaction with the TED. Specifically, in the elastase-cleaved form of ECAM, additional residues (F947 – N963) of the BRD are observed that are disordered in the uncleaved SaA2M structure. All of these residues in the protease cleaved ECAM structure are ordered, owing to the formation of new binding interface between the BRD and the CTMG domain upon protease cleavage (Figure 4-8 a). Critically, the region of the CTMG domain that is involved in BRD binding substantially overlaps with the region of the CTMG domain that forms the binding interface with the TED in the uncleaved form of the protein (Figure 4-8 b). The buried surface area between CTMG and TED in SaA2M is 1972 Å² with 3 hydrogen bonds between these domains whereas the buried surface area between the elastase cleaved BRD and CTMG in ECAM is 1438 Å² with 5 hydrogen bonds between domains.

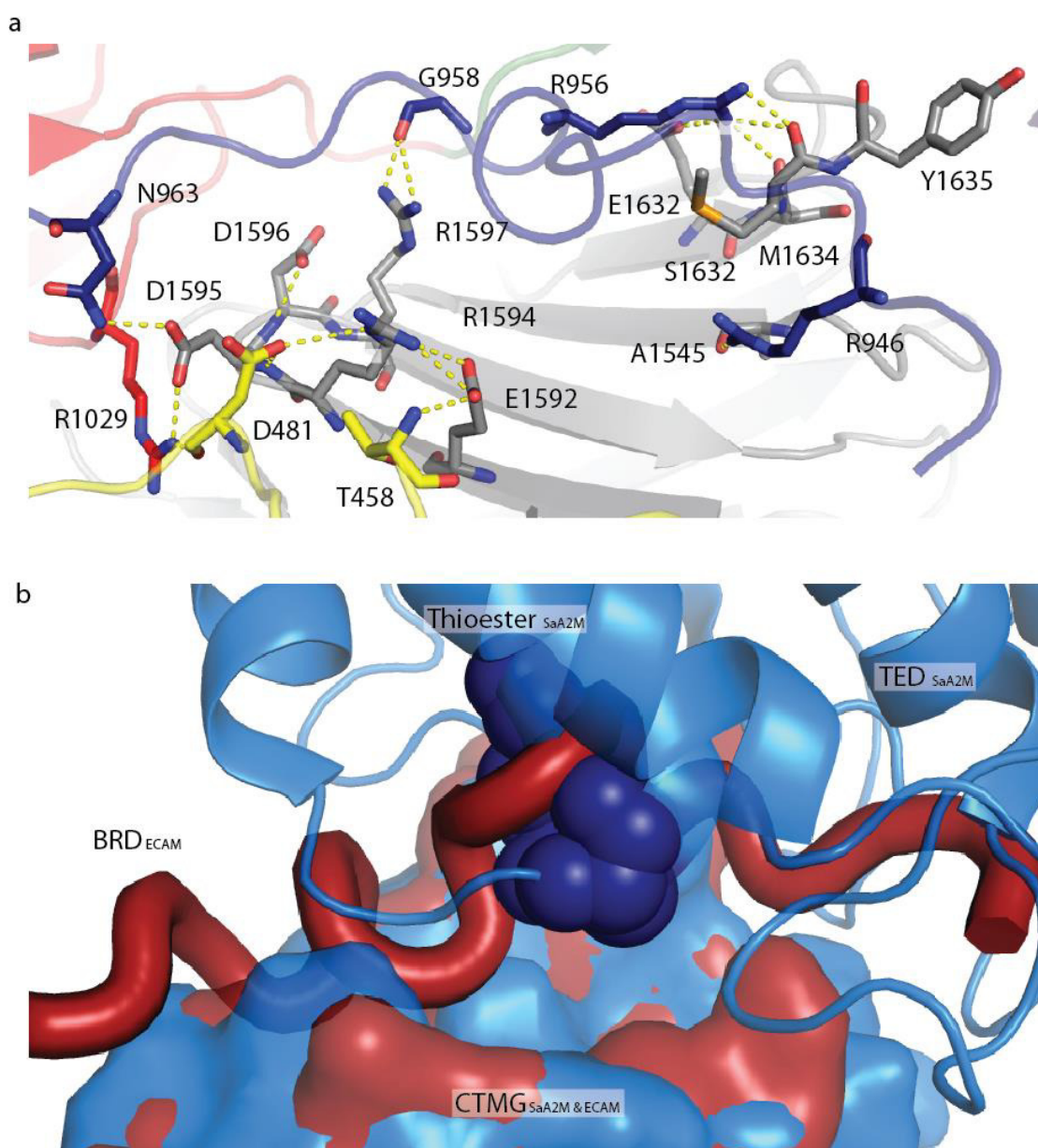


Figure 4-8 Bait region interactions with the CTMG domain.

(a) Hydrogen bonding between the cleaved bait region (blue) and CTMG domain (grey), shown as dashed yellow lines. Note the interaction between R956 and M1634, with the CTMG methionine normally involved in the hydrophobic pocket found in un-cleaved native SaA2M. MG4 and MG9 are shown in yellow and red respectively. (b) Surface representation of the CTMG domain bound to TED in SaA2M (blue) overlaid (by superposition of CTMG domains) with the previously disordered bait region bound to the CTMG domain in elastase cleaved ECAM (red). The thioester bond within the TED is shown using dark blue van der Waals spheres.

In the protease cleaved ECAM, hydrogen bond formation between residues of the CTMG domain and cleaved bait region involves the highly conserved RDDR and EXMY motifs (Figure 4-9). Interestingly, it is residues within these motifs that also form hydrogen bonds with the TED in the uncleaved SaA2M protein (Figure 4-9; ¹⁴).

				TED		CTMG							
				↓		↓							
Escherichia coli A2M (yfhM)	P76578	1181	KAYFYGCLQTA	1192	1592	EFF- DD RFV	1599	1630	MV ES MYVP	1637			
Salmonella typhimurium	Q8ZN46	1173	KAYFYGCLQTT	1184	1583	EFF- DD RFV	1590	1621	QV ES MYAP	1628			
Pseudomonas putida	Q88QC4	1160	QAYFYGCLQTT	1171	1571	EY F - DD RYV	1578	1609	QV ES MYAP	1616			
Bordetella pertussis	Q7VVC2	1222	LT F YFGCTQTI	1233	1663	EFF- DD RYV	1670	1700	E A EDMYRP	1707			
Desulfovibrio alaskensis	Q316S	1188	DRYFYGCLQTA	1199	1582	LQ R DDRFV	1590	1629	SA ED MYDP	1636			
Rhizobium meliloti	Q92VA6	1361	DRYFYGCLQTT	1372	1751	EFF- DD RFV	1758	1791	NV ED MYRP	1798			
Helicobacter hepaticus	Q7VI74	1371	IRYFYGCTQTT	1382	1755	D I R- DD IM	1762	1792	NA E AMYDN	1799			
Fusobacterium nucleatum	D5RAI8	1159	LDYFYICLQIS	1170	1558	D I R- DD VA	1565	1596	KV ES MYNN	1603			
Rickettsia conorii	Q92HD6	1435	DNYFYGCTQLI	1446	1834	NN R - DD VM	1841	1871	YS E AMYDP	1878			
Leptospira interrogans	M5VM24	1485	NR F YSCLEQKL	1496	1863	L T R- EE NWW	1870	1915	RV E AMVLP	1922			

Figure 4-9 Sequence alignments of conserved motifs involved in thioester pocket and conformational activation.

Highlighted in yellow are tyrosines important for maintaining the hydrophobic pocket protecting the thioester bond (CXEQ motif, also shown in yellow). The tyrosine within the C-terminal macroglobulin (CTMG) domain which is also involved in protecting thioester is highlighted in blue as well as it is the glutamate (in CLEQ) that it coordinates with in the native SaA2M structure (PDB: 4U48). The conserved proline found near the thioester is highlighted in green along with the arginine that it coordinates in unactivated SaA2M. The residues highlighted in pink orange and red in the CTMG domain coordinate with residues N963, G958 and R956 respectively, in the cleaved bait region of elastase cleaved ECAM.

4.2.5 BRD cleavage induced conformational shift

To characterise the conformational change seen between the unreacted SaA2M structure and porcine elastase activated ECAM structure these structures were compared by aligning the MG8 domains, which contains the bait region (Figure 4-7 b). Equivalent residues were compared between structures in order to determine the movement in Å shown in table 4.2 with averages represented in figure 4-10. We propose that the loss of the TED interaction with the CTMG domain is sufficient to enable both the large TED movement observed in the cleaved ECAM structure, relative to the unactivated SaA2M structure, and exposure of the thioester bond (Table 4-2, Figure 4-10 a, b, Movie 1). The movement of the TED shows an overall shift of 36 Å, with the MG6 domain moving by 50 Å (Table 4-2, Figure 4-10 a, b, Movie 1). Both the MG6 and TED arms move in to hug the position of where the attacking protease would be located when cleaving the BRD. We also looked at the position of the domains not present in porcine elastase ECAM and noted that if they remained in the position seen in the SaA2M structure, relative to MG8, there would be no clashes with the moving domains (Table 4-2, Figure 4-10 a, b, Movie 1). In addition to these global conformational changes, protease cleavage leads to localised changes in the environment of the thioester bond that likely lead to its activation.

Table 4-2 Movement of domains upon elastase activation

Domain	ECAM Amino acid reference	Distance moved (Å)	SaA2M Amino acid reference
MG4	R398	1	R390
	Q420	1.1	Q412
MG5	M487	14.4	M479
	H581	36.1	H573
MG6	E684	46.4	E676
	A706	57.9	A698
MG8	N985	2.8	N993
	D998	2.1	D1006
MG9	D1048	3.8	D1040
	L1062	10	L1054
CUB	N1153	19.9	N1145
	Q1482	37	Q1473
TED	D1259	39.6	D1251
	Y1311	35.4	Y1303
	L1391	32.5	L1382
	Average of TED	35.8	
MG10	N1540	20.1	R1531
	V1651	14.5	V1642

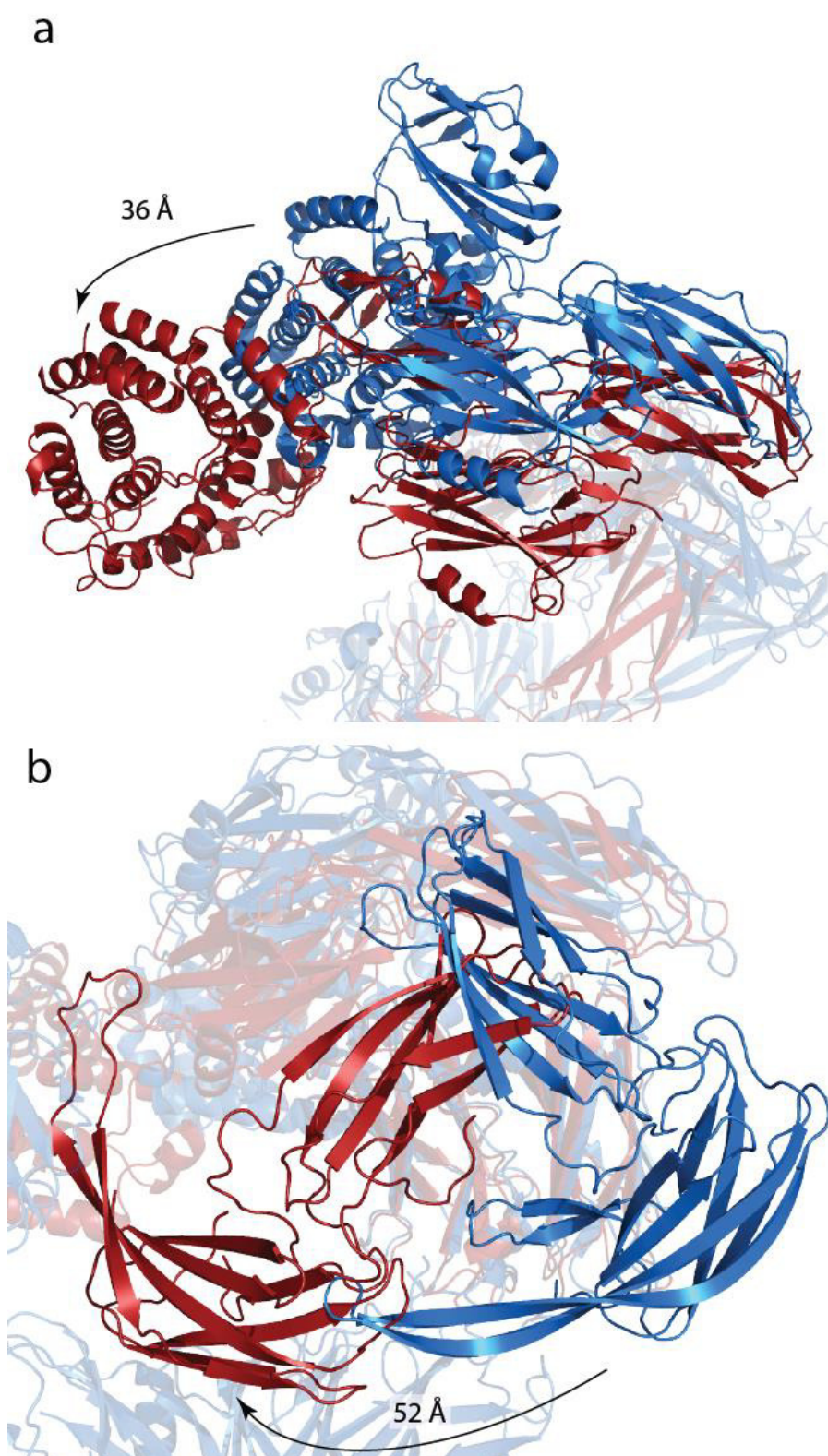


Figure 4-10 Conformational shift of TED relative to MG6 domain of protease-activated ECAM.

(a) Superposition of protease cleaved ECAM (red) and SaA2M (blue) showing 36 Å shift of the TED. (b) Superposition of protease cleaved ECAM (red) and SaA2M (blue) showing 52 Å shift of MG6 domain. Structures overlaid by superposition of MG8 containing the BRD.

4.2.6 Thioester bond activation of α 2Ms

In the unactivated SaA2M structure, the conserved methionine and tyrosine side chains (M1625 and Y1626 in SaA2M) from the CTMG domain form part of the hydrophobic pocket, at the interface with the TED, that has been shown to be important in maintenance of the thioester bond (Figure 4-9, 4-11)¹⁴. The loss of these key side-chain interactions and the movement of the additional sidechains which constitute the protective hydrophobic pocket (Y1175 and Y1177 in SaA2M, and Y1183, Y1185 in ECAM) from the TED expose the thioester bond to the solvent allowing hydrolysis or covalent bond formation with the attacking protease (Figure 4-11). As there are no contacts present from the CTMG domain providing a hydrophobic pocket to protect the thioester from hydrolysis by water, the thioester bond could be cleaved and the deaminated glutamine (Q1190) would be converted to glutamic acid.

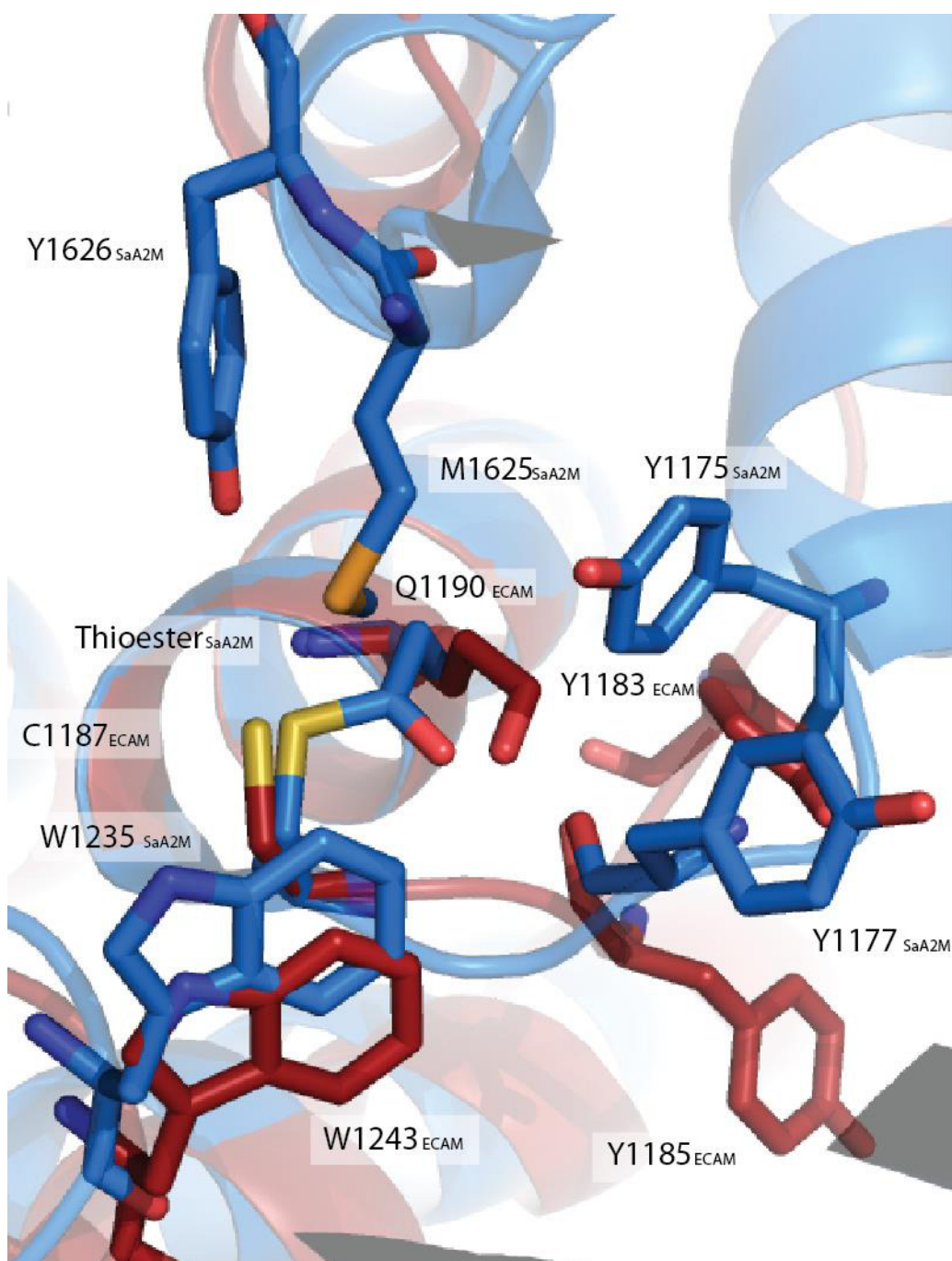


Figure 4-11 Movement of residues within thioester upon protease activation of bacterial $\alpha 2M$.

Overlaid are elastase-cleaved ECAM (red) and native SaA2M (blue) TEDs. Tyrosines thought to protect the thioester region from hydrolysis in native SaA2M (Y1175 and Y1177) are orientated away from the protease-activated ECAM thioester region (Y1183 and Y1185 in protease-activated ECAM).

4.2.7 Biophysical characterisation of ECAM reacted with chymotrypsin

When ECAM is reacted with proteases, a conformational change occurs that may lead to encapsulation and inhibition of the reacting protease. It has been suggested that ECAM functions as an inhibitor of proteases that may include host gut proteases. Within the following work the cleavage of ECAM by the human digestive protease chymotrypsin was investigated. ECAM was reacted with chymotrypsin and purified by SEC with the intention of characterising ECAM in complex with chymotrypsin. Protein from the main peak following reaction of ECAM with chymotrypsin was collected and pooled for AUC and SAXS analysis (Figure 4-12).

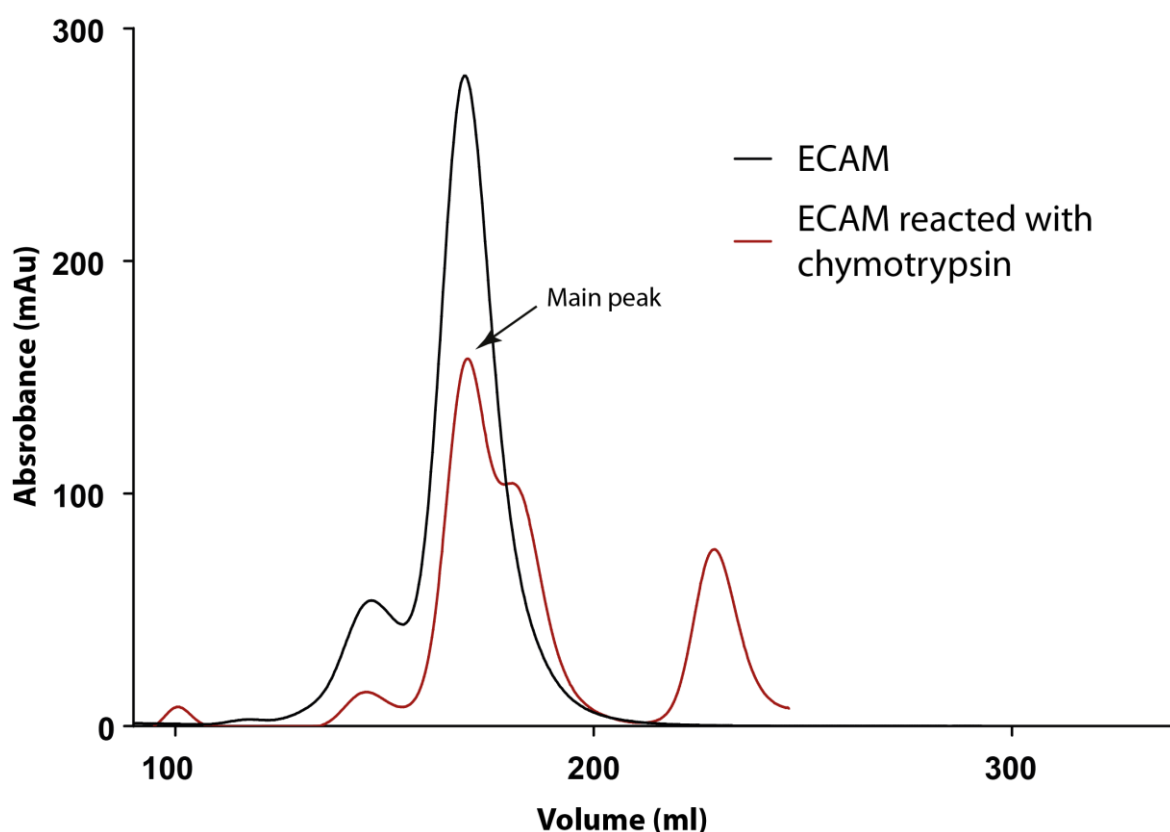


Figure 4-12 Purification by SEC of ECAM reacted with chymotrypsin.

SEC of ECAM that has been reacted with chymotrypsin for 5 minutes on ice and reacted with PMSF protease inhibitor before being loaded onto a Superdex S200 SEC column. The main peak indicated was pooled for analysis by analytical ultracentrifugation and SAXS. ECAM purification shown in black with ECAM reacted with chymotrypsin shown in red.

ECAM reacted with chymotrypsin and purified by SEC was concentrated for analytical ultracentrifugation sedimentation velocity as was previously performed for ECAM. Similar to unactivated ECAM, as the concentration of protease-activated ECAM is increased the concentration of a second species increases with species 2 becoming distinct between 2-4 mg ml⁻¹ (Figure 4-13 a)⁶⁷. The sedimentation coefficient $s_{20,w}^0$ determined for ECAM reacted with chymotrypsin is 5.85 S for species 1 and 8.88 S for species 2 (Figure 4-13 a, b, Table 4-3). When the crystal structure of porcine elastase activated ECAM (PDB:4RTD) is analysed hydrodynamically using SOMO within Ultrascan 2.0, an $s_{20,w}^0$ of 6.68 S is calculated which lies between experimental values for species 1 and 2 of ECAM reacted with chymotrypsin, 5.84 S and 8.88 S respectively¹⁶⁶(Table 4-3). The two species observed are likely monomer and dimer as have been mentioned for ECAM reacted with proteases previously⁶⁷. The calculated sedimentation coefficient of 6.68 S for crystal structure of protease activated ECAM is just under 15% higher than monomer species 1 (5.84 S) indicating that in solution protease reacted ECAM may have more flexibility and elongation than the rigid crystal structure.

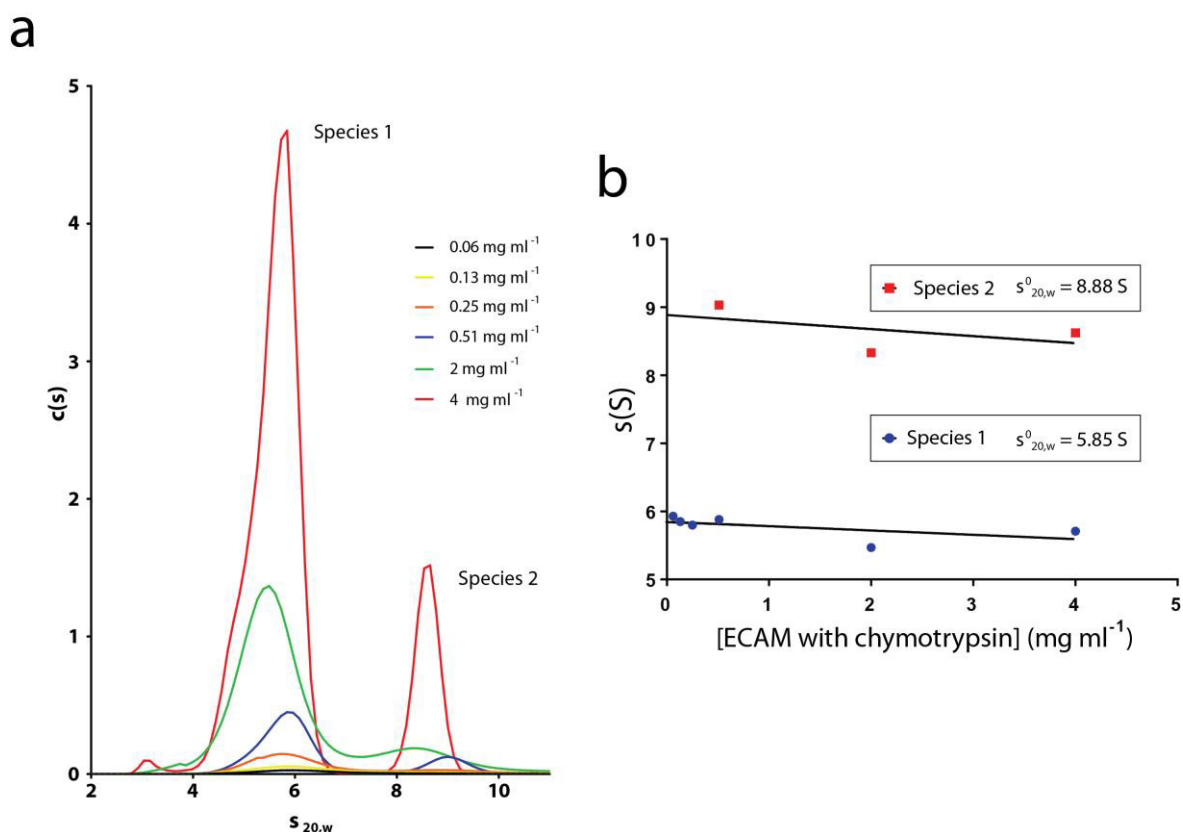


Figure 4-13 AUC analysis of ECAM reacted with chymotrypsin.

(a) Continuous $c(s)$ distribution of ECAM reacted with chymotrypsin at various concentrations. (b) Integrated sedimentation coefficient values at 20°C in water against concentration of ECAM reacted with chymotrypsin.

Table 4-3 Experimental and computed sedimentation coefficient (s) for ECAM reacted with proteases

		ECAM reacted with porcine elastase PDB:4RTD	ECAM reacted with chymotrypsin
$s_{20,w}^0$	Monomer	6.68 S	5.85 S
	Dimer	n/a	8.88 S

4.2.8 Small angle X-ray scattering of ECAM reacted with chymotrypsin and comparison with crystal structure of ECAM reacted with porcine elastase

Dimer formation does not become significant until between 2-4 mg ml⁻¹ and so ECAM reacted with chymotrypsin was analysed by SAXS at concentrations between 0.24 mg ml⁻¹ and 2.3 mg ml⁻¹. The small angle, low concentration data was merged with the wide angle high concentration data for better data analysis and a representative envelope of ECAM reacted with chymotrypsin in solution. The scattering curve for chymotrypsin activated ECAM is representative of a globular protein (Figure 4-14 a). Using the Guinier fit the R_g is calculated as 4.01 nm and reveals no particle interference due to the linearity of the residuals (Figure 4-14 b). Considering the Kratky plot, the protein in solution is globular (Figure 4-14 c). The $p(r)$ plots also show a bell shaped curve indicating a globular protein with a D_{max} of 11.92 nm (Figure 4-14 d).

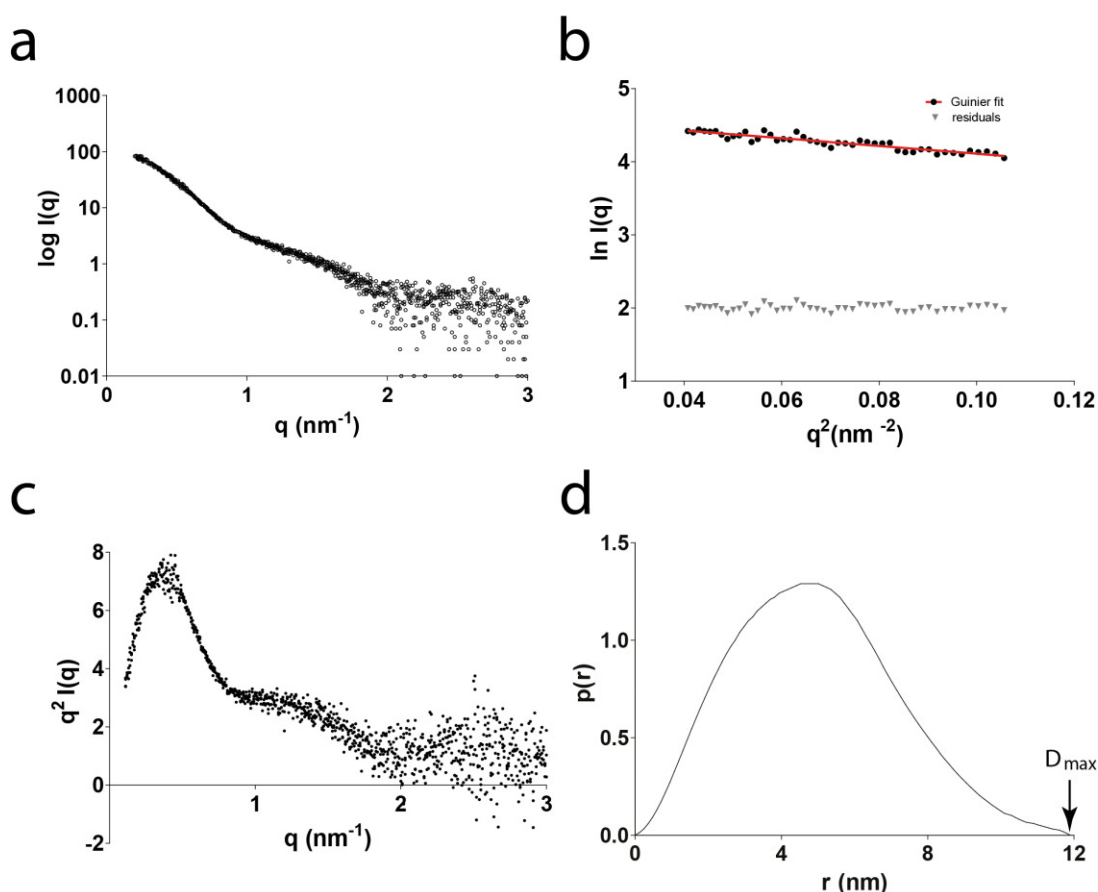


Figure 4-14 SAXS data for ECAM reacted with chymotrypsin.

(a) Raw scattering curve for ECAM (reacted with chymotrypsin) using merged high concentration (2.3 mg ml^{-1}), wide angle and low concentration (0.24 mg ml^{-1}), small angle data. (b) Guinier fit to the raw scattering data in the Guinier region determines an R_g of 4.01 nm. The linearity of the residuals indicates the absence of inter-particle interference. (c) Kratky plot indicates a globular protein. (d) The $p(r)$ distribution analysis indicates a globular protein with a much smaller D_{max} (11.92 nm) as seen for ECAM (18.80 nm).

When the scattering curve of ECAM reacted with chymotrypsin is compared with the crystal structure of ECAM reacted with porcine elastase using CRY SOL, the curves appear very similar at small to wide angles and have a χ value of 0.825 (Figure 4-15)¹⁸⁰. This χ value is a significantly better match to the curve than that seen for either monomer or dimer ECAM (11.867 and 5.471, respectively). The I_0 calculated molecular weight for chymotrypsin reacted ECAM is 102 kDa and does not match the molecular weight of the porcine elastase reacted molecular weight of 127 kDa. The SOMO simulated R_g and D_{max} values for porcine elastase reacted ECAM of 3.8 nm and 12.42 nm are close to the

experimental values seen for chymotrypsin reacted ECAM of 4.01 nm and 11.92 nm (Table 4-4)¹⁶⁶. However, the larger D_{\max} seen for the crystal structure indicates that the chymotrypsin sample in solution is slightly more compact (Table 4-4). The close CRY SOL fit and similarity in R_g and D_{\max} values seen between experimental and simulated crystal structure confirm that the protein sample in solution is representative of the crystal structure.

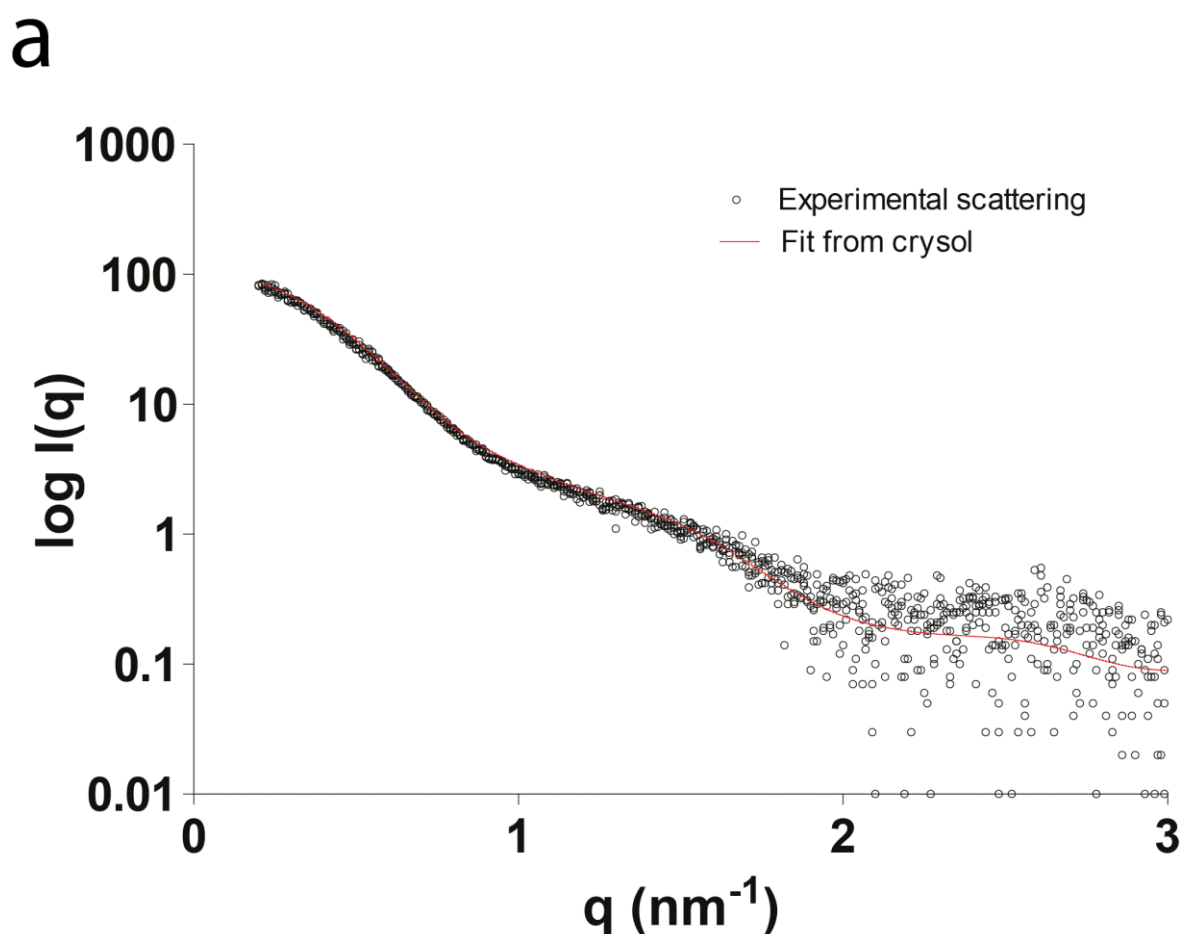


Figure 4-15 CRY SOL comparison between elastase activated ECAM crystal structure and chymotrypsin activated ECAM SAXS curve.

(a) Chymotrypsin activated ECAM (Black, merged wide angle data at 2.3 mg ml^{-1} and small angle data at 0.24 mg ml^{-1}) comparison with porcine elastase activated ECAM crystal structure (red, PDB: 4RTD) follows scattering data. SAXS data comparisons with crystal structures performed using CRY SOL with a χ value of 0.825^{180} .

Table 4-4 Experimentally determined and computed ECAM biophysical properties

	Chymotrypsin treated ECAM	Elastase treated ECAM crystal PDB:4RTD
MW (kDa)	102 [I_0]	127
Rg (Guinier) (nm)	3.9	3.8
D_{\max} (nm)	11.92	12.42

The low resolution envelope generated using DAMMIF for ECAM reacted with chymotrypsin has a very similar shape when compared to the crystal structure of ECAM reacted with porcine elastase (Figure 4-16)¹⁷⁹.

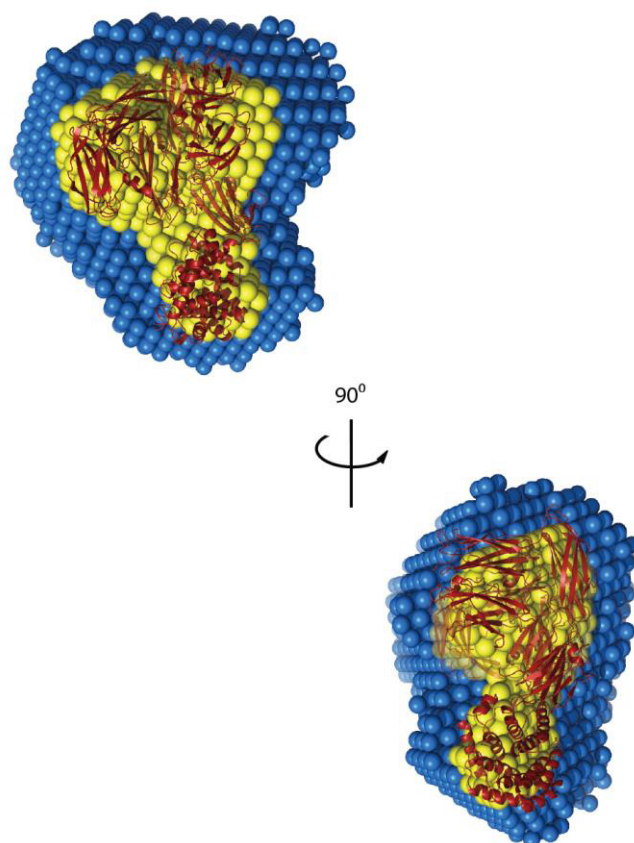


Figure 4-16 Crystal structure of elastase treated ECAM overlaid with SAXS envelope of chymotrypsin treated ECAM.

Elastase treated ECAM crystal structure (PDB: 4RTD) aligned within SAXS envelope generated using DAMIF with SAXS data collected from Chymotrypsin treated ECAM (merged wide angle data at 2.3 mg ml^{-1} and small angle data at 0.24 mg ml^{-1}). DAMIFILT envelope shown in yellow and DAMAVER envelope in blue¹⁷⁶. Models aligned manually in Pymol.

4.3 Discussion

Although it has been suggested that small molecule activated human α 2M resembles the conformation of the protease cleaved form, there has to date been no detailed structural data to confirm this ²⁵. The similarity in overall conformation of protease cleaved ECAM with methylamine activated human α 2M confirms that the protease cleaved form and chemically activated form of human α 2M are structurally equivalent. In addition, these data suggest that the entrapment of cleaving proteases likely occurs in a similar manner for BA2Ms as has been proposed for human α 2M ^{13,188}. However, there are clearly key differences in the details of the interactions that maintain the inactive conformation of BA2Ms and human α 2M, since the recent structures of unactivated and methylamine activated SaA2M show that chemical cleavage of the thioester bond does not in this case lead to global conformational changes ^{13,14}. This difference is likely due to the domain location of the side-chains that comprise the thioester protecting pocket (Figure 4-17). For BA2Ms, this pocket comprises of 2 tyrosine side chains from the TED and a tyrosine and a methionine side chain from the CTMG domain whereas in eukaryotic α 2Ms all four residues (3 tyrosine side chains and one methionine side chain) are found in the CTMG domain. When SaA2M is reacted with methylamine no conformational change is seen, but Y1175 from the TED is displaced ¹⁴. The structural counterpart of this side chain (Y1307 from TEP1) in eukaryotic α 2M family members is, however, found within the CTMG domain (Figure 4-17)¹⁰.

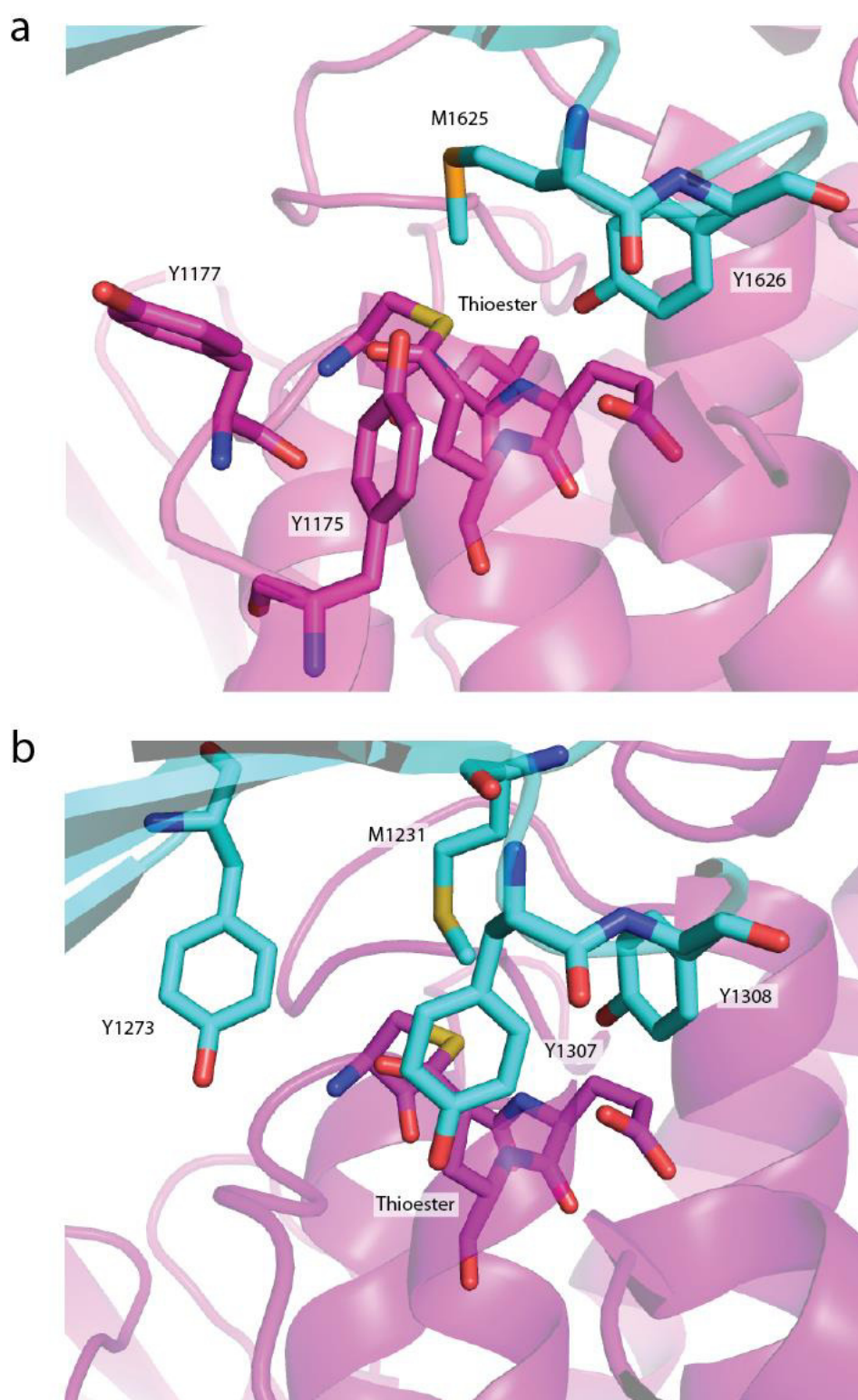


Figure 4-17 Comparison of the thioester protecting pocket in eukaryotic and bacterial α 2Ms.

The thioester pocket highlighting conserved residues from the TED (magenta) and CTMG domain (cyan) of (a) SaA2M and (b) the eukaryotic α 2M family member TEP1 (PDB 4D94; Le et al., 2012). Highlighted side chains are conserved among bacterial and eukaryotic α 2Ms, respectively.

It is presumably the rearrangement of this side chain and perhaps other CTMG side chains that comprise the thioester protecting pocket, which leads to loss of TED-CTMG domain binding and subsequent global conformational changes. The buried surface area seen between TED-CTMG in native SaA2M is 27% higher than the surface area buried on formation of the BRD-CTMG in protease cleaved ECAM, however, the number of hydrogen bonds increases from 3 to 5 (Figure 4-8). Although there is a decrease in the buried surface area, the movement in residues involved in the hydrophobic pocket protecting the thioester is what triggers conformational change releasing TED allowing subsequent interaction with a cleaving protease.

The role of BA2Ms, which are inner membrane anchored periplasmic proteins, has been suggested as protease inhibitors that inhibit exogenous proteases that have breached the outer membrane. During the course of this work a crystal structure of trypsin activated ECAM was published¹⁸⁷. This crystal structure showed distinct similarity to porcine elastase activated ECAM although all domains were present minus the first N-terminal MG domain (Figure 4-18 a). With an r.m.s.d between porcine elastase activated ECAM and trypsin activated ECAM of 0.711 using 761 α -carbons, these two forms of protease-activated ECAM display very similar structures regardless. The interaction of the cleaved bait region with CTMG is also observed in the trypsin activated ECAM structure although not analysed within their paper (Figure 4-18 b). Importantly the BRD R956 is also seen to hydrogen bond with M1634 confirming the interaction seen within the elastase activated crystal structure. In the trypsin activated structure, trypsin was present within the crystal lattice, however, not forming a repeating pattern that could diffract. Another difference noted was a slightly rotated TED that could be due to this domain being bound to a protease molecule in trypsin activated ECAM (Figure 4-18 a).

Investigation of ECAM reacted with chymotrypsin by SAXS resulted in a scattering curve that closely matches the CRY SOL simulated curve for ECAM reacted with porcine elastase (Figures 4-14, 4-15, 4-16, Table 4-3, 4-4). This indicates that regardless of the identity of the cleaving protease the conformational changes in ECAM are similar and closely resemble those in HA2M.

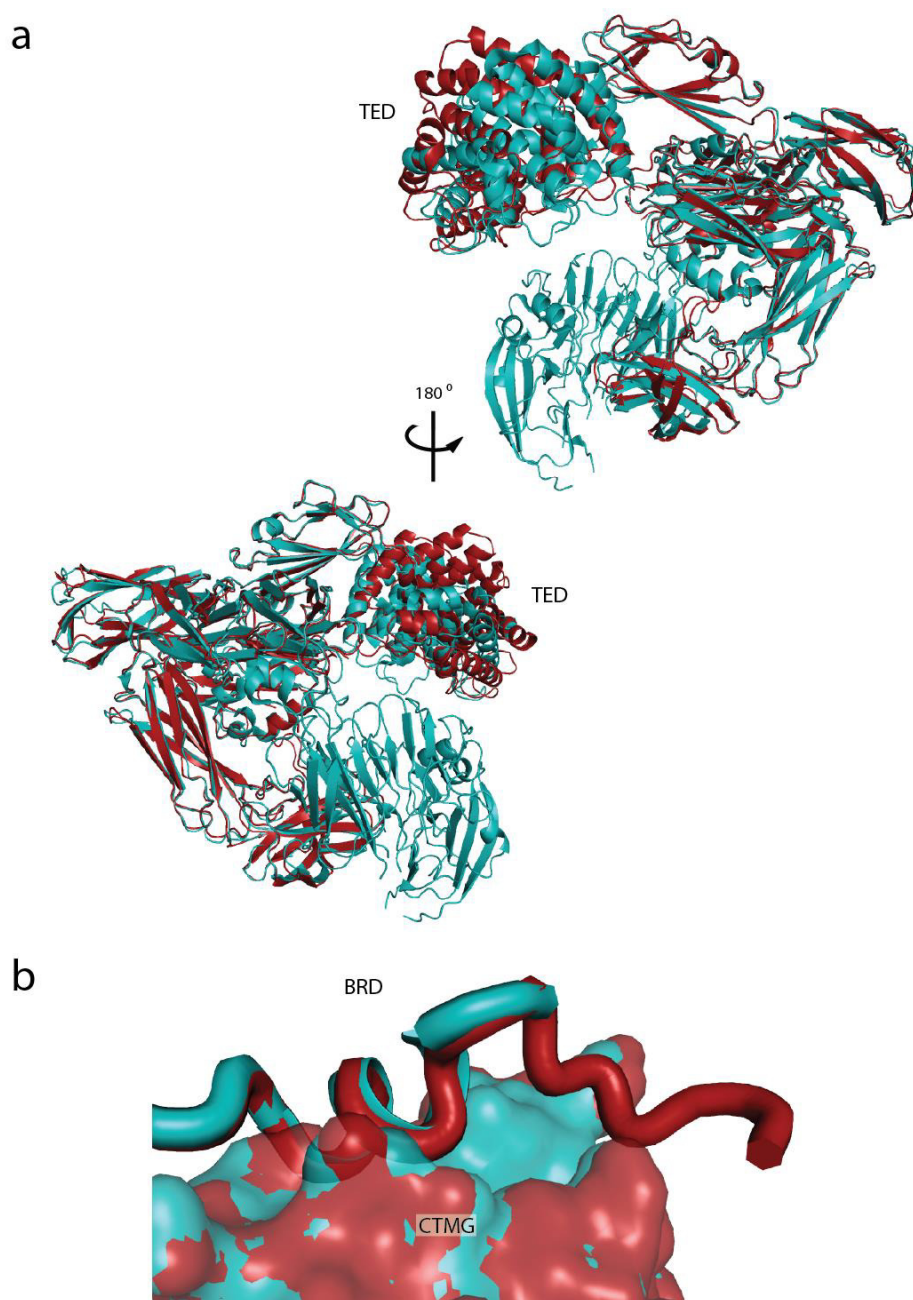


Figure 4-18 Alignment of protease-activated ECAM structures.

The alignment of porcine elastase activated ECAM (Red, PDB:4RTD) and trypsin activated ECAM (Cyan, PDB:4ZIQ)¹⁸⁷. The full structures shown in (a) have a very similar orientation with an r.m.s.d of 0.711 (using 761 α -carbons) between structures however, a slight difference in orientation of the TED and extra domains for the trypsin activated ECAM are observed. (b)Trypsin activated ECAM displays similar interactions between the cleaved bait region and the CTMG as porcine elastase activated ECAM. In trypsin activated ECAM a BRD is observed with a slightly shorter length however maintaining the important residue contacts seen in porcine elastase activated ECAM.

Comparison of the structures of unactivated SaA2M and protease cleaved ECAM illustrates how protease induced conformational changes may enable protease entrapment (Figure 4-14).

The movement of the TED arm and MG6 domain arm around the central pocket above the bait region is similar to the movement seen in methylamine activated human $\alpha 2M$. When accounting for the domains not present in the protease-activated ECAM structure, no clashes are seen between the moving arms and the absent domains ¹³. The entrapment of proteases would limit the proteolysis of smaller substrates as has been suggested for human $\alpha 2M$ and would prevent the cleavage of important larger substrates such as the peptide component of the peptidoglycan layer or large proteins (Figure 4-19) ¹⁶.

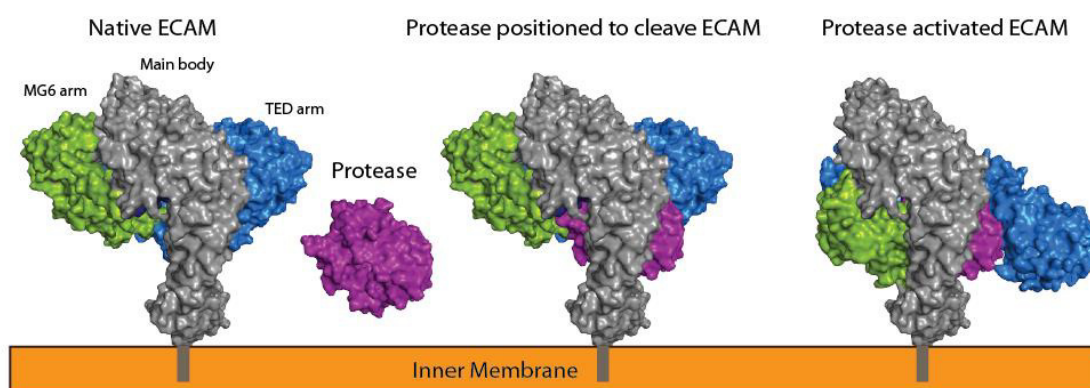


Figure 4-19 Putative mechanism of protease entrapment and inhibition by ECAM.

ECAM with membrane anchor within the periplasm of bacteria encounters protease and forms a covalent complex upon cleavage, inhibiting the protease from cleaving large substrates. The main body of ECAM anchored to the inner membrane is shown in grey and contains the domains showing little movement (MG4, 8 & BRD) between native and protease-activated structures or that are not present in the protease-activated structure (MG1-3 & 7). The blue TED arm (TED, CUB, CTMG & MG9 domains) and green MG6 arm (MG5 & 6 domains) entrap the protease with the blue arm containing the thioester that forms a covalent bond.

Although we cannot be sure why the covalently bound elastase is not present in our structure, this may be due to a lack of available and correctly positioned lysine side chains on the surface of the protease, since the thioester bond is preferentially cleaved by this side chain¹³. The lack of MG domains 1, 2, 3 and 7 within the crystal lattice is likely due to the MG domains being cleaved by elastase. With both the native SaA2M structure and the trypsin activated structure showing a dimer there is likely a physiological importance to this.

The structure of protease-activated ECAM suggests a competitive mechanism of activation in which cleavage of the BRD allows the normally intrinsically disordered region of this domain to outcompete the TED for CTMG domain binding. Loss of the TED-CTMG domain interaction leads to a large conformational rearrangement of ECAM and exposure of the reactive thioester bond. The structural similarity between methylamine activated human α 2M and protease cleaved ECAM suggests that similar mechanisms likely operate across the diverse members of the α 2M family.

5 The role of Penicillin-binding protein 1C in the function of ECAM

Circular dichroism data was collected and processed by Sharon Kelly.

5.1 Introduction

Escherichia coli possess two genes encoding bacterial α 2M type proteins. As discussed in previous chapters ECAM, one of these two proteins, has been characterised previously^{6,67} (Figure 5-1). The gene encoding ECAM (*yfhM*) exists in an operon with *pbpC* which encodes the nonessential penicillin binding protein Pbp1C (Figure 5-1). It has been suggested that these proteins constitute a defence and repair mechanism, with ECAM inhibiting invading host proteases and Pbp1C repairing any proteolytic damage to the peptidoglycan^{6,67,123}.

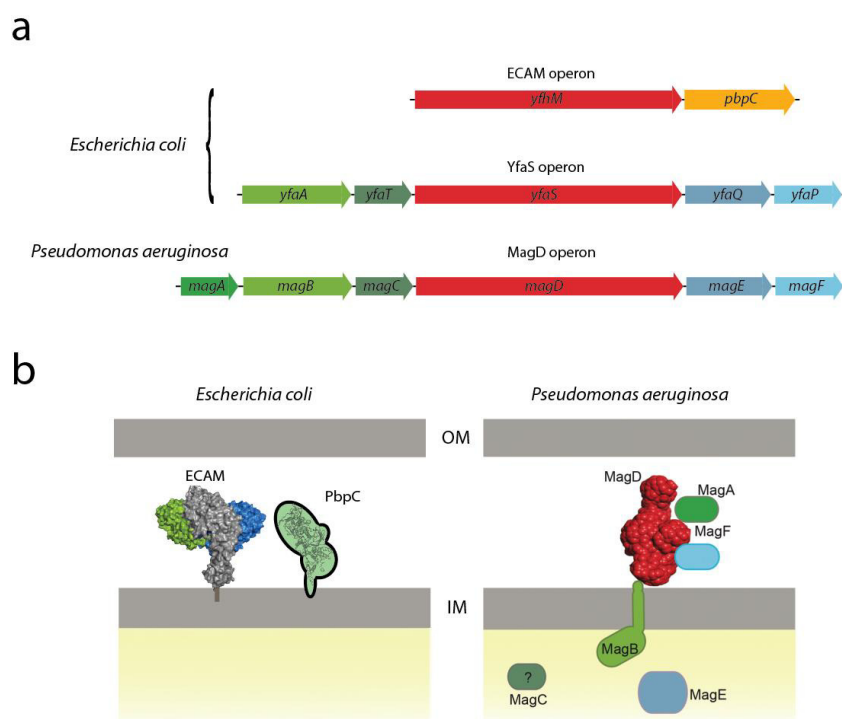


Figure 5-1 Bacterial α -2-macroglobulin operons and a schematic representation of their expressed proteins localized to the bacterial envelope.

(a) Bacterial α 2M operons ECAM and YfaS are found in *Escherichia coli* and YfaS equivalent operon MagD is found in *Pseudomonas aeruginosa*. (b) Within *E. coli*, ECAM and Pbp1C are both translocated to the periplasm across the inner membrane. MagD in *P. aeruginosa*, is translocated across the inner membrane to the periplasm and is associated with the inner membrane, possibly through interactions with MagB, which harbours one transmembrane domain in its C-terminus. The orientation of the MagB in the inner membrane was chosen arbitrarily. MagA and MagF, encoded by the same operon, are also partners of the membrane associated Mag complex. Figure 5-1 b reproduced from Robert-Genthon *et al.* 2013⁶⁸.

Pbp1C has been shown to have transglycosylase (TGase) function however although a putative transpeptidase (TPase) domain has been suggested no TPase activity has been observed¹²³. Previous characterisation of Pbp1C was performed on protein purified from membranes and sepharose columns covalently bound with Pbp1C retained Pbp1B, Pbp2,3¹²³. Although Pbp1C is not essential and cannot replace murein production in [Pbp1A, Pbp1B] double knockouts, deletion of Pbp1C results in an altered production of murein¹²³. Compared to wildtype *E. coli*, the knockout of *pbpC* resulted in a decrease in peptidoglycan synthesis by 75%¹²³. There was also noted a change in peptidoglycan composition with a 371% increase in dimeric subunits that were crosslinked between a tetrapeptide disaccharide and a pentapeptide disaccharide¹²³. This was suggested as occurring due to Pbp1Cs functioning as a complex with other peptidoglycan synthesis proteins such as Mgt and Pbp1B¹²³.

Within the genome of *E. coli* K12 there is a second operon including a gene encoding a bacterial α 2M, *yfaS*. This operon also contains further genes, *yfaA*, *yfaT*, *yfaQ* and *yfaP*, with no characterisation of the encoded proteins performed to date. *Pseudomonas aeruginosa* contains only one copy of bacterial α 2M, similar to *E. coli yfaS*, called MagD (Figure 5-1). The MagD operon contains all of the proteins found in the *E. coli yfaS* operon and an additional gene upstream of the homologue *magB* (Figure 5-1).

The family of penicillin binding proteins includes several enzymes with varied function, with class A having both transglycosylase and transpeptidase function, class B having transpeptidase function and class C having carboxypeptidase or endopeptidase function. In this chapter *in vivo* characterisation of ECAM and Pbp1C through western blotting and the purification and characterisation of Pbp1C is described. This work was undertaken in order to gain insight into the possible function of Pbp1C both individually and in concert with its operon partner ECAM.

5.2 Results

5.2.1 *In vivo* analysis of ECAM

ECAM is a member of the $\alpha 2M$ superfamily and has been suggested to inhibit proteases in a similar fashion to $\alpha 2M$ ^{14,67,77}. To test this hypothesis the behaviour of ECAM *in vivo* was investigated. Western blotting was performed using anti-ECAM antibodies on whole cell lysate. When wild type (WT) *E. coli* K12 cell extracts were blotted using the anti-ECAM antibody five distinct bands were observed at 260, 180, 160, 120 and 100 kDa (Figure 5-2). The bands at 260 and 180 kDa could be dimer and monomer species respectively, as seen for purified ECAM (Figure 5-2, 3-1 c).

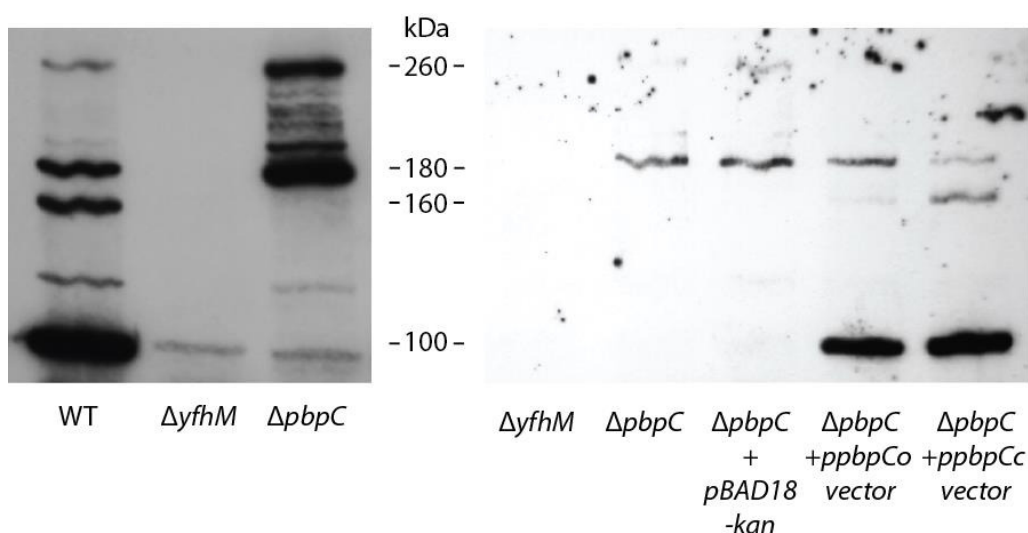


Figure 5-2 Wild-type phenotype loses ECAM cleavage upon *pbpC* knockout and returns upon complementation.

Wildtype (WT) *E. coli* K12 has four significant bands present at 260, 180, 160, 120 and 100 kDa. The ECAM knockout strain, $\Delta yfhM$, has lost all four major bands with only a faint band present at 100 kDa. *E. coli* K12 with *pbpC* knocked out ($\Delta pbpC$) has no band at 160 kDa and has lost the major band at 100 kDa with an increase in intensity in bands seen at 260 and 180 kDa. $\Delta pbp1C$ has further bands at various molecular weights between 160 and 180 kDa. Bands seen in wildtype at 160 and 100 kDa and that are absent in *pbp1c* knockout are restored on complementation with plasmid based *pbpC* but remain absent in the presence of pBAD18-kan vector control. Cell extracts were analysed by Western blotting using anti-ECAM antibodies.

Since the most intense band is at 100 kDa this suggests that ECAM is present mostly in a cleaved form in growing cells. One possibility is that there is processing of ECAM by unidentified bacterial proteins. Since *yfhM* is generally linked to the gene encoding Pbp1C (*pbpC*), the effect of loss of Pbp1C on processing of ECAM was considered. Interestingly in a $\Delta pbpC$ strain the intense band (at 100 kDa) representing cleaved ECAM is absent and bands at 180 and 260 kDa show increased intensity (Figure 5-2). In addition, a number of bands between 180 and 260 kDa are also apparent that are absent in WT *E. coli*. Thus Pbp1C is required for processing of ECAM although from these experiments it is not possible to determine if this requires the proteins to directly interact. The laddering effect seen between 260 and 180 kDa in the $\Delta pbpC$ strain may be due to ECAM covalently attaching to proteins in close proximity upon heat denaturation prior to SDS-PAGE (Figure 5-2).

In order to confirm that Pbp1C is involved in ECAM processing, the gene encoding Pbp1C including the lipid anchor was cloned into the pBAD18-kan vector and used to complement the $\Delta pbpC$ strain. Two plasmids were generated, one without a stop codon in order to subsequently add a Myc-tag (ppbpCo), the second with a stop codon (ppbpCc). Both complementation plasmids were successful in returning the $\Delta pbpC$ strain to the WT phenotype with the most intense ECAM band observed at 100 kDa (Figure 5-2).

5.2.2 Role of other Pbps on ECAM processing

To determine if ECAM processing is specific to Pbp1C, Western blotting with anti-ECAM antibodies of cell lysates from strains lacking other penicillin binding proteins and proteins involved in peptidoglycan synthesis was performed. In strains lacking the class A penicillin binding proteins Pbp1A and Pbp1B ($\Delta mrcA$ and $\Delta mrcB$ respectively) processing of ECAM was similar to WT with an intense ECAM band observed at 100 kDa by western blotting with anti ECAM antibodies (Figure 5-3). The knockout for the gene encoding the EPase Pbp7 ($\Delta pbpG$) as well as knockouts for genes encoding CPases Pbp4, Pbp5 and Pbp6 ($\Delta dacA$, $\Delta dacB$ and $\Delta dacC$ respectively) also display a wild type phenotype (Figure 5-3). The knockouts for genes encoding a β -lactamase, the monofunctional TGase and a periplasmic protease ($\Delta ampC$, $\Delta opgH$ and Δprc respectively) all maintain a wild type phenotype with 160 and 100 kDa ECAM cleavage products present (Figure 5-3).

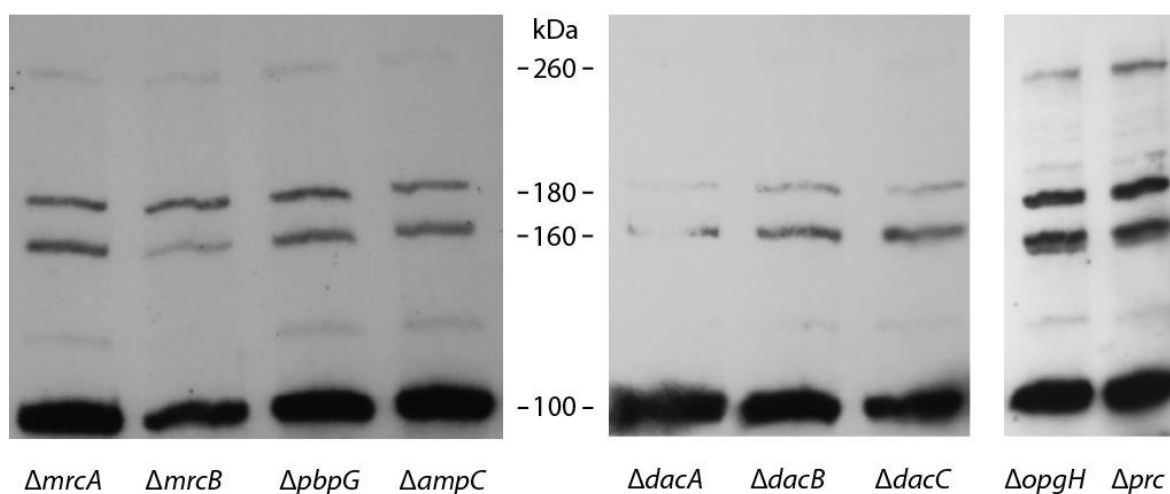


Figure 5-3 Strains lacking genes involved in the synthesis of peptidoglycan maintain a wildtype ECAM processing phenotype.

Class A Penicillin binding protein knockouts $\Delta mrcA$ and $\Delta mrcB$ retain the 160 and 100 kDa bands seen in WT *E. coli*. The WT phenotype is also maintained across various peptidase penicillin binding protein knockouts as well as the monofunctional murein synthase (*opgH*) and *prc* that encodes a periplasmic protease. Cell extracts were analysed by Western blotting using anti-ECAM antibodies.

Class C penicillin binding proteins include endopeptidases such as Pbp7 and carboxypeptidases such as Pbp4. The domain that is inhibited by penicillin in these Pbps is very similar in structure to the domain found in class A and class B Pbps. When β -lactams such as ampicillin react with Pbps they covalently link to the active site residue of the peptidase domain and inactivate catalytic function be it transpeptidase, carboxypeptidase or endopeptidase. Our hypothesis was that by adding ampicillin at a concentration just above the minimum inhibitory concentration Pbp1C would be chemically 'knocked out' resulting in a similar phenotype as seen in the $\Delta pbpC$ strain, with respect to ECAM processing (Figure 5-3). On addition of ampicillin there is an increase in intensity of bands at 180 and 260 kDa however the cleavage product bands do not disappear indicating partial inhibition of Pbp1C function (Figure 5-3).

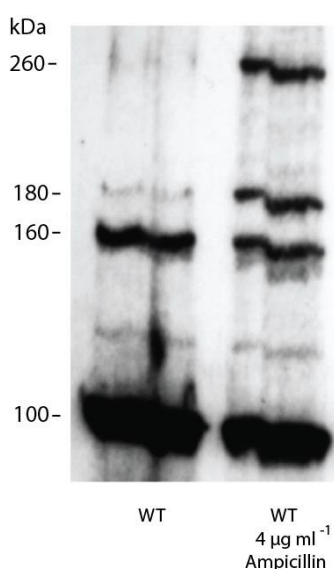


Figure 5-4 Chemical inhibition of Pbp1C in WT shows similar phenotypic characteristics seen in $\Delta pbpC$

Bands at 260 and 180 kDa appear when WT is grown in the presence of a minimum inhibitory concentration of ampicillin (4 $\mu\text{g ml}^{-1}$) indicating less cleavage of ECAM and a build-up of monomer and dimer as seen in $\Delta pbpC$. Cell extracts were analysed by Western blotting using anti-ECAM antibodies.

5.2.3 *Pbp1C* expression, refolding and purification

In order to determine if Pbp1C directly interacts with ECAM and is involved in ECAM processing I attempted to purify Pbp1C. The gene encoding Pbp1C, *pbpC*, was cloned including an N-terminal 6-histidine tag. The gene amplified by PCR was cloned into pET21a using NdeI and XhoI. This plasmid was transformed into BL21(DE3) *E. coli* cells for inducible expression. Cells were grown at 37°C in LB, to an OD₆₀₀ of 0.6 and 0.1 mM IPTG was added to induce Pbp1C expression. Cells were grown for a further four hours and lysed. Although Pbp1C was being expressed at high levels these were not soluble and even with modification in levels of IPTG and growth temperature, the protein remained insoluble as a white inclusion body. Several gradual N-terminal domain truncations were cloned in order to attempt to obtain a soluble Pbp1C however these all resulted in similar inclusion bodies.

At this time I showed that ampicillin inactivates Pbp1C, therefore I decided to use kanamycin as a selection marker (Figure 5-4). The *pbpC* gene was cloned into pET28a which uses kanamycin for selection. However, the protein remained insoluble, forming inclusion bodies under various expression conditions.

It was therefore decided to attempt to solubilise, refold and purify Pbp1C from inclusion bodies expressed from the plasmid pCF4, consisting of full-length *pbpC* within the pET28a kanamycin vector. The inclusion body was washed using low concentrations of detergent and solubilised in 5% sodium lauroyl sarcosinate (sarkosyl), 50 mM Tris-HCl, 200mM NaCl, pH 7.5. This solubilised protein was then dialysed into 0.5% Triton X100 50 mM Tris-HCl, 200mM NaCl, pH 7.5 at 4°C overnight. This protocol resulted in soluble PbpC with a good level of purity. However, due to the various problems associated with Triton X100 such as significant absorbance at the wavelength used to measure protein concentration (280 nm), several other detergents were tested. The refolding protocol was performed with TWEEN 20, β OG (n-Octyl β -D-Glucopyranoside), CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), DDM (n-Dodecyl β -D-

maltoside), deoxycholic acid and LDAO (Lauryldimethylamine oxide) used at the respective critical micelle concentrations for dialysis in order to find an alternative detergent to triton X100. Detergent solubilised Pbp1C samples were run on an analytical S200 size exclusion column to compare how these detergents performed in refolding Pbp1C. Equal weights of inclusion body were used with each detergent in order to compare refolding yield. The detergents CHAPS, deoxycholic acid and TWEEN 20 all resulted in multiple peaks and so were not considered further for refolding. DDM had a very low yield and so was also not considered further. The two best candidates for refolding Pbp1C were LDAO and BOG. LDAO resulted in a sharp peak and was used for future experiments (Figure 5-5).

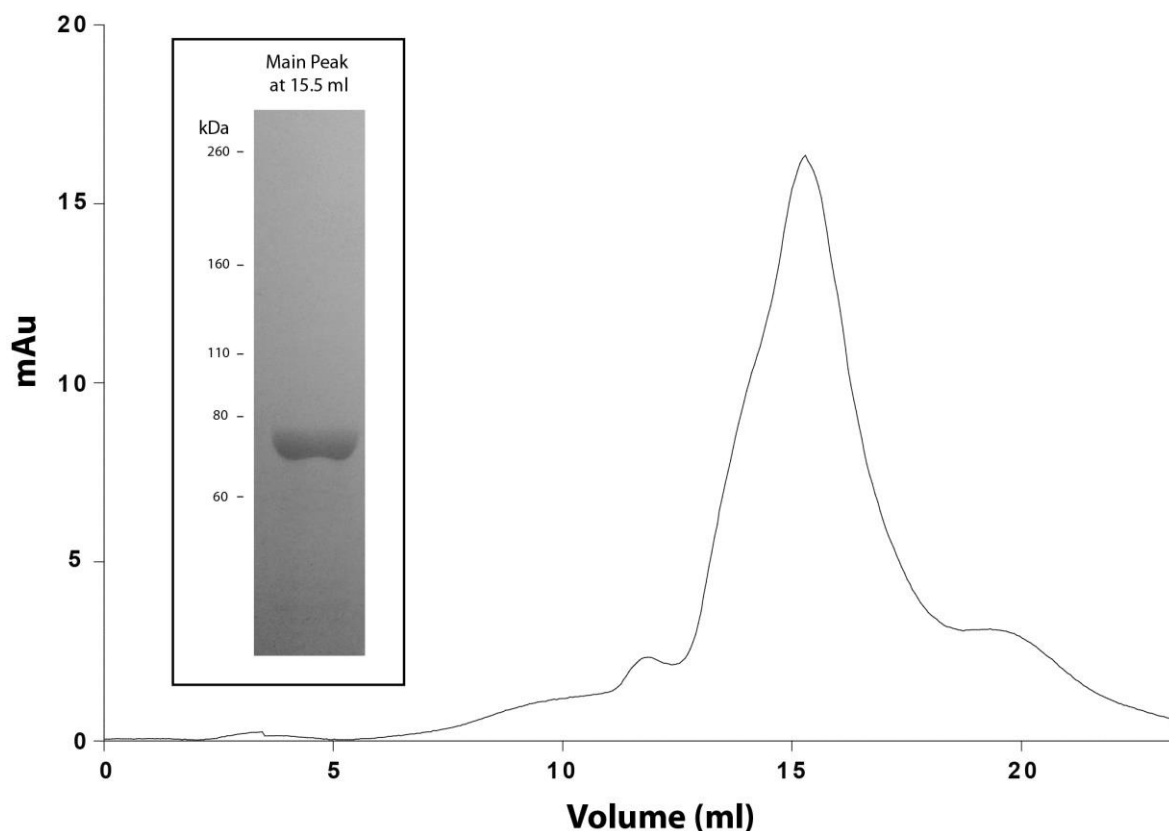
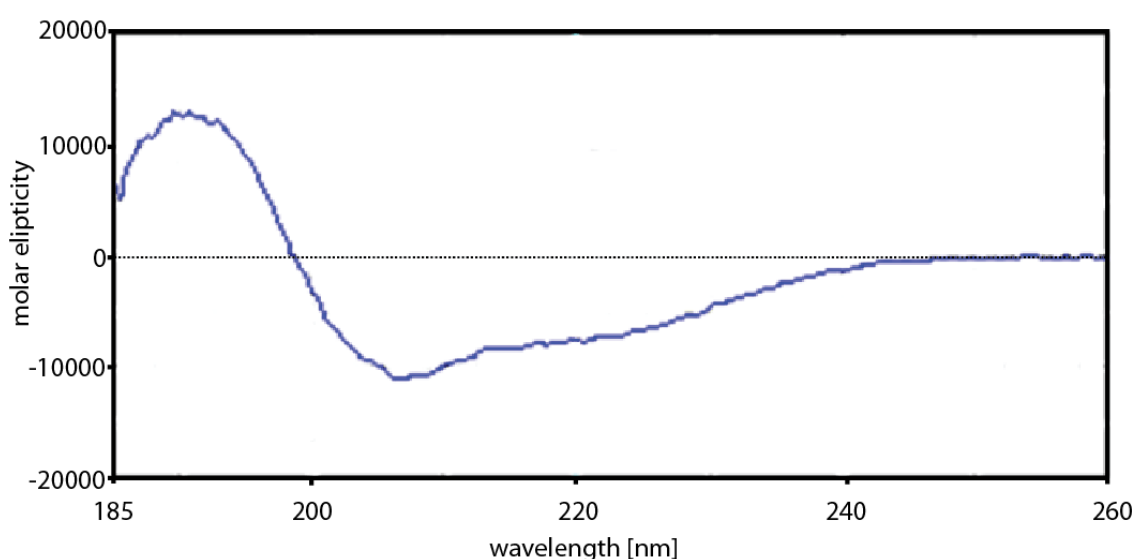


Figure 5-5 SEC of LDAO refolded Pbp1C.

A small analytical S200 column was used to purify Pbp1C refolded using various detergents. Shown is the chromatogram of Pbp1C refolded using LDAO (Lauryldimethylamine oxide). Pbp1C runs by SDS-PAGE at a molecular weight of approximately 75 kDa.

In order to further confirm that Pbp1C had been successfully refolded, circular dichroism spectroscopy was performed to estimate secondary structure present using the Provencher & Glockner method (Figure 5-6 a)¹⁶². YASPIN and Jpred online secondary structure prediction servers were used to predict the secondary structure for Pbp1C (Figure 5-6 b)^{189,190}. The calculated secondary structure based on CD is 26.6% α -helix and 25.3% β -sheet. This is in good agreement with the predicted secondary structure of PbpC from YASPIN and Jpred.

a



b

Secondary Structure prediction method		α -helix (%)	β -strand (%)	Other (%)
Provencher & Glockner method from CD	Result 1	25.9	21.8	52.4
	Result 2	27.4	28.7	43.8
	Average	26.6	25.3	48.1
YASPIN		33.6	18.3	48.0
Jpred		29.1	12.2	58.7

Figure 5-6 Circular dichroism spectroscopy indicates that Pbp1C has native secondary structure.

(a) Far-UV spectrum of Pbp1C (6 μ M) in 0.05 M sodium phosphate, 0.1% LDAO, pH 7.5 shows that greater than 50% of the protein has secondary structure elements. Estimates are based on the average of all matching solutions and calculated using the Provencher & Glockner method¹⁶². (b) Table showing the results from CD calculated secondary structure elements compared to the primary sequence predicted secondary structure using the YASPIN and Jpred online servers^{189,190}.

Further structural characterisation of Pbp1C that had been refolded with LDAO was attempted by SAXS however due to the presence of detergent micelles the data was not suitable for analysis. Any attempt to dialyse the LDAO out resulted in aggregation of the protein. Crystal screens were also set up at 10 and 20 mg ml⁻¹ Pbp1C using the Molecular Dimensions crystallisation screens MORPHEUS®, MIDAS™, JCSG-plus, and PACT premier however the few crystals that appeared were salt.

As Pbp1C was successfully refolded it was decided to see if Pbp1C is directly responsible for the cleavage of ECAM observed *in vivo* by Western blotting. To test this, 2.5 µM of Pbp1C was reacted with 2.5 µM of ECAM in 50 mM Tris, 200 mM NaCl, pH 7.5 with 0.1% LDAO at room temperature for 15 minutes. This reaction resulted in no apparent cleavage of ECAM indicating that there is no cleavage of ECAM by Pbp1C under these conditions (Figure 5-7). This lack of cleavage may be due to the fact Pbp1C may not have been refolded into an enzymatically active state. The experiment was repeated including a final concentration of 1 mM of magnesium, calcium or potassium within the buffer, however no cleavage was observed.

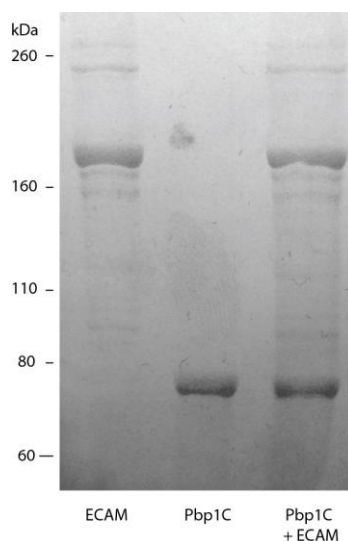


Figure 5-7 Refolded Pbp1C shows no processing of ECAM *in vitro*

The combined sample of ECAM with Pbp1C shows the same composition as the proteins individually with no cleavage products.

5.3 Discussion

The work discussed in this chapter sought to determine if Pbp1C is involved in the function of ECAM. Interestingly, deletion of *pbpC* results in a significant change in ECAM processing, such that cleaved ECAM species that are present in WT *E. coli* are absent in a $\Delta pbpC$ strain. Instead the bands representing full length and dimeric ECAM, show increased intensity. The western blots are performed on denatured samples and so any dimers would be expected to denature and appear as a single band, however due to the presence of a thioester able to form a covalent linkage, denatured ECAM could be covalently binding to a second molecule of ECAM (Figure 5-2). Alternatively, it may be that during denaturation of the dimer form seen in solution, this becomes entangled resulting in the observed dimer by SDS-PAGE. The covalent binding of ECAM to other proteins upon denaturation could also explain the laddering seen between 260 kDa and 180 kDa (Figure 5-2). The various bands seen between the 180 kDa monomer and 260 kDa dimer upon deletion of *pbpC* could be proteins that form a complex with ECAM and Pbp1C under normal conditions (Figure 5-2). These proteins, in close proximity to native ECAM become covalently bound upon denaturation of ECAM.

The bands seen at 160 kDa and 100 kDa within wildtype cells are cleavage products due to the processing of ECAM *in vivo* and the bait region is where ECAM is cleaved by proteases *in vitro* (Figure 5-2)¹⁴. Upon denaturation prior to SDS-PAGE this cleavage would result in two fragments of ECAM, the N-terminus to bait region domain and BRD to C-terminus (100 kDa and 80 kDa respectively). The bands seen by Western blot using anti-ECAM antibodies of wild type cells resulted in major bands at 260 kDa, 180 kDa, 160 kDa and 100 kDa. As mentioned previously the 260 kDa band could be a dimer of ECAM covalently bound via the thioester during denaturation and the 180 kDa band a monomer of ECAM. There remains a question of what the 160 kDa and 100 kDa bands consist of? From our data showing the knockout of the gene encoding Pbp1C results in the loss of cleavage products of ECAM it could be the case that Pbp1C is directly responsible for the processing of ECAM. As such, the 160 kDa band could represent Pbp1C (calculated 85 kDa by sequence, 75 kDa by SDS-PAGE) covalently bound to the 80 kDa BRD to C-terminus

fragment of ECAM (Figure 5.8). The 100 kDa band could represent the N-terminus to BRD fragment upon cleavage at the bait region or may be a protease (approximately 20 kDa in size) that cleaves ECAM and has been covalently bound to the BRD to C-terminus fragment (Figure 5.8).

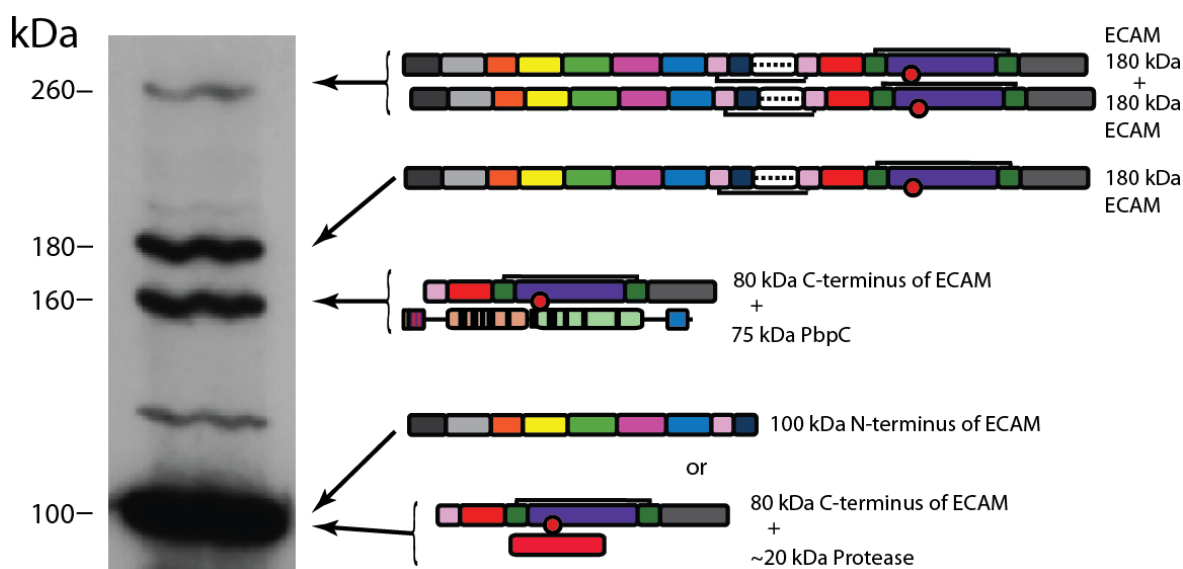


Figure 5-8 Possible composition of bands seen by Western blot.

The band seen at 260 kDa may be a dimer of two ECAM molecules that have covalently bound by thioester bond upon denaturation as they are in close proximity natively. The band seen at 180 kDa is full length ECAM. Pbp1C may cleave ECAM at the bait region the 80 kDa C-terminus that contains the thioester covalently binds Pbp1C resulting in a 160 kDa complex. As the protein sample has been denatured the 100 kDa N-terminus of ECAM separates or may be a complex between an unknown protease (~20 kDa in size) covalently bound to the C terminal fragment of ECAM.

The possible interaction of ECAM and Pbp1C may constitute a regulation mechanism, which through steric hindrance limits access to the Pbp1C 'transpeptidase' active site. Instead of Pbp1C being a transpeptidase, crosslinking sidechains within the peptidoglycan network, the peptidase domain may cleave peptide sidechains similar to a LD- or DD-CPase. This enzyme could have specificity for the bait region of ECAM allowing conformational activation encapsulating Pbp1C, limiting the cleavage of side chains to the subunits Pbp1C has incorporated *via* its transglycosylase domain (Figure 5-9). The peptides found at the terminus of the pentapeptide subunit are D-amino acids whereas the peptides within ECAM are L-amino acids suggesting that if Pbp1C is cleaving ECAM it

would not have enzyme activity for peptidoglycan pentapeptides, however further work is required to investigate this.

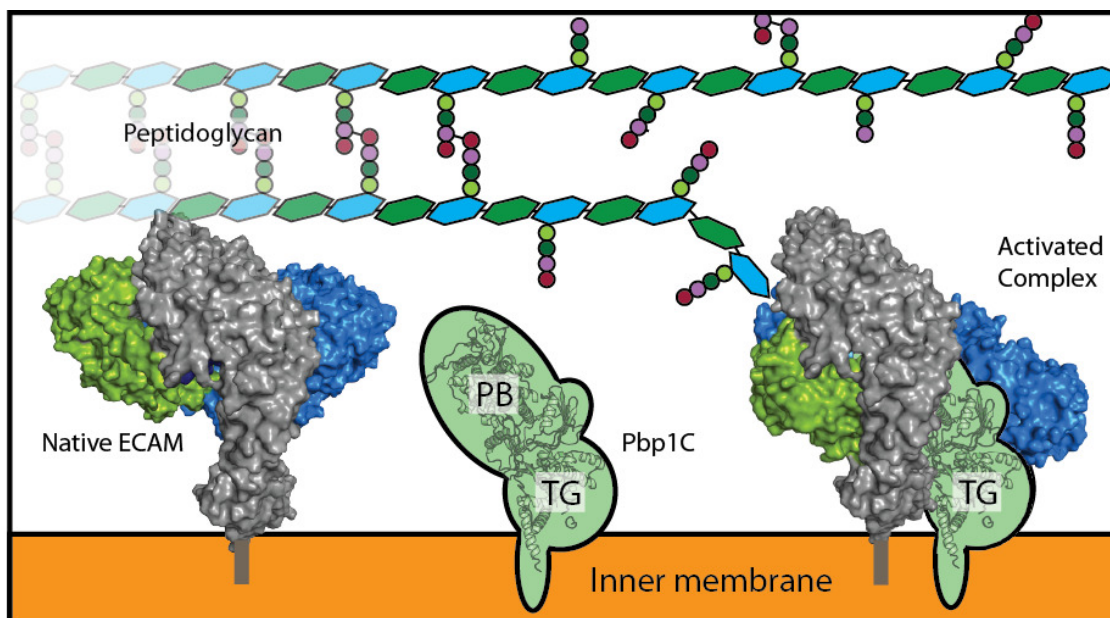


Figure 5-9 Model of how Pbp1C and ECAM function in live cells.

Pbp1C functioning as a LD- or DD-carboxypeptidase cleaves ECAM resulting in a complex that limits the function of Pbp1C to cleave peptides from subunits transported by its transglycosylase (TGase) domain.

The refolding of Pbp1C appears to have been successful as CD data show secondary structure comparable with that predicted by secondary structure (Figure 5-6). The lack of any cleavage of ECAM by refolded Pbp1C however, if cleavage is directly mediated by Pbp1C, could indicate that the active site is not functional (Figure 5-7). The requirement of detergent in order to keep Pbp1C stable is most likely due to the transglycosylase domain normally interacting with the membrane. The lack of any cleavage of ECAM by refolded Pbp1C adds evidence against its direct involvement in ECAM processing, and further work is required to confirm or disprove the hypothesis that Pbp1C directly cleaves ECAM. As the peptidase domain may be involved with the cleavage of peptides from peptidoglycan, it may be possible to purify this isolated domain as a soluble protein to directly test this hypothesis.

6 Concluding remarks

One branch of the $\alpha 2M$ superfamily are protease inhibitors that act by covalently linking to the attacking protease and sterically hindering its ability to bind substrate. ECAM is a member of this branch of $\alpha 2M$ found in *E. coli* that has been predicted to form a defence and repair mechanism alongside Pbp1C, a protein involved in peptidoglycan synthesis^{6,67}.

Using analytical ultracentrifugation sedimentation velocity we have shown that unreacted ECAM exists in a monomer-dimer equilibrium (Figure 3-2 b, 3-3, 3-4, 3-5). Both the ECAM monomer and dimer are more elongated in solution when compared to the native crystal model of SaA2M, a *Salmonella enterica* homologue of ECAM (Table 3-2). Further investigation of the monomer and dimer of ECAM by SAXS confirms the presence of a monomer and dimer species in solution that is more elongated than the crystal dimer of SaA2M (Table 3-2).

There remains a question of why ECAM exists in a monomer-dimer equilibrium within the periplasm. The bait region within the dimer high resolution model of SaA2M is buried, forming an interface with the MG domains (MG1 and MG2) involved in binding to the membrane (Figure 6-1 a). Since in the SaA2M dimer structure these MG domains block access to the bait region the dimer form is most likely an inactive form, perhaps acting as a reservoir available to replace cleaved monomeric ECAM (Figure 6-1 b).

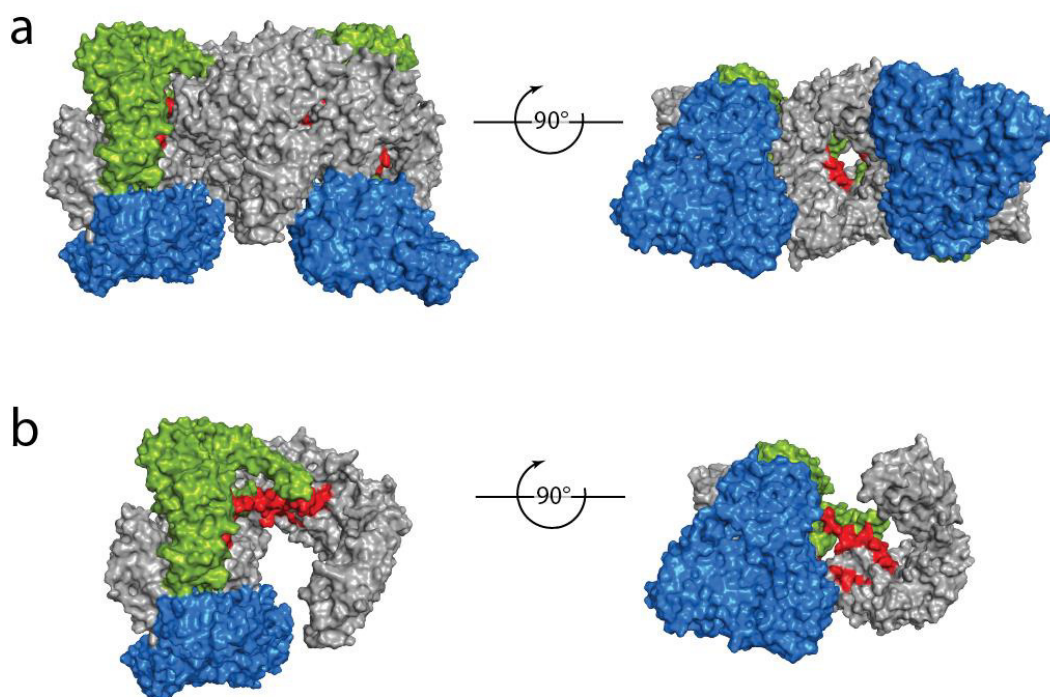


Figure 6-1 Availability of bait region for cleaving proteases.

Surface representation of SaA2M as dimer with main MG body coloured grey, MG6 arm coloured green, TED arm coloured blue and bait region coloured red. (a) SaA2M dimer showing the limited availability of the bait region in the dimer. (b) SaA2M monomer with a bait region clear of obstruction. (PDB: 4U48¹⁴)

Although attempts to crystallise a native unreacted form of ECAM were unsuccessful, crystals of elastase cleaved ECAM were obtained (Figure 4-2 a). This protease-activated form not only matches the domain orientation of the small molecule activated form of HA2M but also has an almost identical domain arrangement to trypsin activated ECAM with only a small amount of movement apparent by the TED (Figure 4-7 a, 4-18 a). Although no evidence could be found that crystals of porcine elastase activated ECAM contained elastase within the lattice, trypsin activated ECAM contains trypsin within its crystal lattice although the protease is not ordered enough to give strong diffraction¹⁸⁷. The difference in orientation of the TED between trypsin and elastase activated structures may be due to trypsin being bound to this domain. SAXS analysis of ECAM cleaved by chymotrypsin resulted in a low resolution envelope that appears very similar to the high resolution model of ECAM reacted with porcine elastase (Figure 4-15, Table 4-4). Further,

the CRYSOLO curve generated for porcine elastase reacted ECAM closely resembles the experimental scattering curve for ECAM reacted with chymotrypsin (Figure 4-14)¹⁸⁰. This shows that upon protease cleavage a similar conformational change occurs regardless of the protease that cleaves ECAM.

In porcine elastase cleaved ECAM, the cleaved bait region interacts with the CTMG. Importantly this interaction includes a residue that normally interacts with the TED to form a hydrophobic pocket that protects the thioester from solvent (Figure 4-17). The interaction of the cleaved bait region with the CTMG is also seen in trypsin activated ECAM¹⁸⁷. The presence of this interaction between CTMG and cleaved bait region in both high resolution models of cleaved ECAM (trypsin and porcine elastase cleaved ECAM) adds further evidence that the disruption of the TED-CTMG interaction is the trigger for conformational activation in members of the α 2M family (Figure 4-18 b).

Previous work has suggested that ECAM functions alongside Pbp1C as a protease inhibitor that defends against host proteases breaching the outer-membrane with Pbp1C repairing any damage to the peptidoglycan^{6,67}. In this work it was shown that Pbp1C is involved in ECAM processing, although a direct effect could not be demonstrated *in vitro*. We show that in wild type cells cleavage of ECAM occurs in the absence of exogenous proteases (Figure 5-2). The cleavage products seen in wildtype ECAM are not observed in a strain lacking *pbpC* but return upon complementation of the Δ *pbpC* with plasmid based *pbpC* (Figure 5-2). The loss of these breakdown products in this deletion mutant could indicate that Pbp1C is involved in the regulation of ECAM. The lack of any interaction seen upon addition of ECAM to refolded Pbp1C suggests that the relationship is indirect; however, the lack of interaction may be due to incorrect/incomplete refolding of Pbp1C. It could also be suggested that Pbp1C cleaves ECAM *in vivo* at the bait region forming a complex between the two (Figure 5-8). The band seen at 160 kDa could be the complex between Pbp1C and the C-terminal ECAM fragment (that contains a thioester covalently linking both) and the band at 100 kDa could be the remaining N-terminal fragment (Figure 5-8). Similar to α 2M, ECAM has been shown to inhibit proteases through covalently binding to the cleaving protease and restricting the proteases access to larger substrates, however

allowing access to small substrates⁶⁷. If the hypothesis that ECAM forms a complex with Pbp1C is correct then ECAM may act as a regulator of Pbp1C function, limiting access of certain substrates to the active site (Figure 6-2). Although the suggestion of Pbp1Cs cleavage ECAM could explain the observed processing of ECAM further work is required to confirm if this is the case (Figure 6-2). To confirm that Pbp1C is able to cleave ECAM, mutagenesis on the active sites of the transpeptidase and transglycosylase domains could be performed.

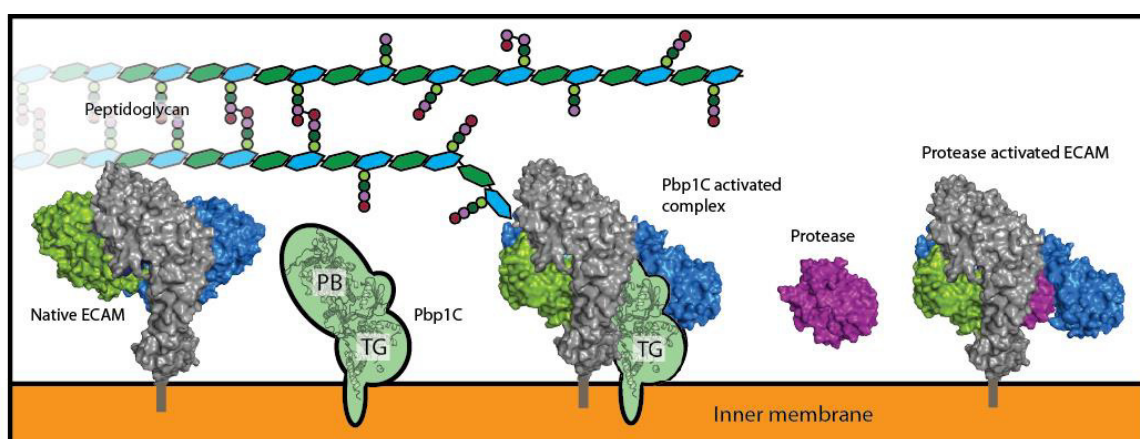


Figure 6-2 Model of ECAM reacting with Pbp1C and proteases.

Native monomer ECAM available for cleavage could react with Pbp1C limiting its function to small subunits or inhibition of invading protease.

7 References

1. Barrett, A. J. & Starkey, P. M. The Interaction of α -2-Macroglobulin with Proteinases. *J. Biochem.* **133**, 709–724 (1973).
2. With, C., Westergaard, J. G. & Sottrup-jensensli, L. Characterization of Human Pregnancy Zone Protein. *J. Biol. Chem.* **260**, 15723–15735 (1985).
3. Lin, M. Cell surface antigen CD109 is a novel member of the α 2 macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood* **99**, 1683–1691 (2002).
4. Li, Z.-F., Wu, X. & Engvall, E. Identification and characterization of CPAMD8, a novel member of the complement 3/ α 2-macroglobulin family with a C-terminal Kazal domain. *Genomics* **83**, 1083–93 (2004).
5. Galliano, M.-F. *et al.* A novel protease inhibitor of the α 2-macroglobulin family expressed in the human epidermis. *J. Biol. Chem.* **281**, 5780–9 (2006).
6. Budd, A. *et al.* Bacterial α 2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5**, R38 (2004).
7. de Bruijn, M. H. & Fey, G. H. Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 708–712 (1985).
8. Blanchong, C. a. *et al.* Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. *Int. Immunopharmacol.* **1**, 365–392 (2001).
9. Tack, B. F. *et al.* Fifth component of human complement: purification from plasma and polypeptide chain structure. *Biochemistry* **18**, 1490–1497 (1979).
10. Baxter, R. H. G. *et al.* Structural basis for conserved complement factor-like function in the antimalarial protein TEP1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11615–11620 (2007).
11. Sottrup-Jensen, L. *et al.* Common evolutionary origin of α 2-macroglobulin and complement components C3 and C4. *Proc Natl Acad Sci USA* **82**, 9–13 (1985).
12. Dodds, A. W. & Law, S. K. A. The phylogeny and evolution of the thioester bond-containing proteins. *Immunol. Rev.* **166**, 15–26 (1998).
13. Marrero, A. *et al.* The crystal structure of human α 2-macroglobulin reveals a unique molecular cage. *Angew. Chemie - Int. Ed.* **51**, 3340–3344 (2012).
14. Wong, S. G. & Dessen, A. Structure of a bacterial α 2-macroglobulin reveals mimicry of eukaryotic innate immunity. *Nat. Commun.* **5**, 4917 (2014).
15. Le, B. V., Williams, M. *et al.* Molecular Basis for Genetic Resistance of *Anopheles gambiae* to *Plasmodium*: Structural Analysis of TEP1 Susceptible and Resistant Alleles. *PLoS Pathog.* **8**, (2012).
16. Sottrup-Jensen, L. α -Macroglobulins: Structure, Shape, and Mechanism of Proteinase Complex Formation. *J. Biol. Chem.* **264**, 11539–11542 (1989).
17. G. S. Salvesen & J. Travis. Human plasma proteinase inhibitors. *Biochem. Annu. Rev.* **110**, 256–261 (1983).

18. Cohn, E. J. & Strong, L. E. Preparation and properties of serum and plasma proteins; a system for the separation into fractions of the protein and lipoprotein components of biological tissues and fluids. *J. Am. Chem. Soc.* **68**, 459–475 (1946).
19. Cassiman, J. & Berghe, H. Van Den. Human Pregnancy Zone Protein and alpha-2-Macroglobulin. *J. Biol. Chem.* **261**, 16622–16625 (1986).
20. Barrett, B. A. J. *et al.* The Electrophoretically 'Slow' and 'Fast' Forms of the α_2 -Macroglobulin Molecule. *Biochem. J.* **181**, 401–418 (1979).
21. Enghild, J. J. *et al.* Proteinase binding and inhibition by the monomeric α -macroglobulin rat α_1 -inhibitor-3. *J. Biol. Chem.* **264**, 11428–11435 (1989).
22. Sottrup-Jensen, L. *et al.* The α -Macroglobulin Bait Region. *J. Biol. Chem.* **264**, 15781–15789 (1989).
23. Sottrup-Jensen, L. *et al.* A THIOL-ESTER IN α_2 -MACROGLOBULIN CLEAVED DURING PROTEINASE COMPLEX FORMATION. *FEBS Lett.* **121**, 275–279 (1980).
24. Armstrong, P. B. & Quigley, J. P. α_2 -macroglobulin : an evolutionarily conserved arm of the innate immune system. *Dev. Comp. Immunol.* **23**, 375–390 (1999).
25. Tapon-breaudiere, J. *et al.* Electron microscopy of the conformational changes of alpha 2-macroglobulin from human plasma. *EMBO J.* **4**, 85–89 (1985).
26. Huang, W. *et al.* P. G. Localization of basic residues required for receptor binding to the single alpha-helix of the receptor binding domain of human alpha2-macroglobulin. *Protein Sci.* **7**, 2602–2612 (1998).
27. Raymond, W. W. *et al.* alpha(2)-Macroglobulin Capture Allows Detection of Mast Cell Chymase in Serum and Creates a Reservoir of Angiotensin II-Generating Activity. *J. Immunol.* **182**, 5770–5777 (2009).
28. WICHER, V. & DOLOVICH, J. *et al.* INTERACTIONS OF BACILLUS-SUBTILIS ALKALINE PROTEINASES WITH α_2 -MACROGLOBULIN AND α_1 -ANTITRYPSIN. *Int. Arch. Allergy Appl. Immunol.* **40**, 779–780 (1971).
29. Peslova, G. *et al.* Hepcidin, the hormone of iron metabolism, is bound specifically to alpha-2-macroglobulin in blood. *Blood* **113**, 6225–6236 (2009).
30. Sarma, J. V. & Ward, P. a. The Compliment System. *NIH Public Access author Manuscr.* **343**, 227–235 (2012).
31. Hee, C. K. *et al.* Regenerative Tendon and Ligament Healing: Opportunities with Recombinant Human Platelet-Derived Growth Factor BB-Homodimer. *TISSUE Eng. PART B-REVIEWS* **18**, 225–234 (2012).
32. Bramanti, V. *et al.* Astroglial-conditioned media and growth factors modulate proliferation and differentiation of astrocytes in primary culture. *Neurochem. Res.* **32**, 49–56 (2007).
33. Nancey, S. *et al.* Serum interleukin-6, soluble interleukin-6 receptor and Crohn's disease activity. *Dig. Dis. Sci.* **53**, 242–247 (2008).
34. Wiesmann, C. *et al.* Structure of C3b in complex with CR1g gives insights into regulation of complement activation. *Nature* **444**, 217–220 (2006).

35. Bork, P. & Beckmann, G. The CUB Domain A widespread module in developmentally regulated proteins. *J. Mol. Biol.* **231**, 539–545 (1993).
36. Doan, N. & Gettins, P. G. W. Human alpha2-macroglobulin is composed of multiple domains, as predicted by homology with complement component C3. *Biochem. J.* **407**, 23–30 (2007).
37. Law, S. K. & Dodds, a W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263–274 (1997).
38. Ikai, A. *et al.* A recombinant bait region mutant of human α 2-macroglobulin exhibiting an altered proteinase-inhibiting spectrum. *Cytotechnology* **31**, 53–60 (1999).
39. Van Rompaey, L. *et al.* Design of a new protease inhibitor by the manipulation of the bait region of a-2-macroglobulin: inhibition of the tobacco etch virus protease by mutant a-2-macroglobulin. *Biochem. J.* **195**, 191–195 (1995).
40. Hoffmann, J. A. *et al.* R. A. B. Phylogenetic perspectives in innate immunity. *Science (80-.).* **284**, 1313–1318 (1999).
41. Medzhitov, R. & Janeway, C. Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* **173**, 89–97 (2000).
42. Walport, M. J. Advances in immunology: Complement (First of two parts). *N. Engl. J. Med.* **344**, 1058–1066 (2001).
43. Walport, M. J. Advances in immunology: Complement (Second of two parts). *N. Engl. J. Med.* **344**, 1140–1144 (2001).
44. Janssen, B. J. C. *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511 (2005).
45. Kidmose, R. T. *et al.* Structural basis for activation of the complement system by component C4 cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15425–15430 (2012).
46. Ricklin, D. & Lambris, J. D. Complement-targeted therapeutics. *Nat. Biotechnol.* **25**, 1265–1275 (2007).
47. Mizuno, Y. *et al.* ISOLATION AND CHARACTERIZATION OF A MANNAN-BINDING PROTEIN FROM RAT-LIVER. *J. Biol. Chem.* **256**, 4247–4252 (1981).
48. Drickamer, K. *et al.* L. MANNOSE-BINDING PROTEINS ISOLATED FROM RAT-LIVER CONTAIN CARBOHYDRATE-RECOGNITION DOMAINS LINKED TO COLLAGENOUS TAILS - COMPLETE PRIMARY STRUCTURES AND HOMOLGY WITH PULMONARY SURFACTANT APOPROTEIN. *J. Biol. Chem.* **261**, 6878–6887 (1986).
49. Ezekowitz, R. A. B. *et al.* A HUMAN MANNOSE-BINDING PROTEIN IS AN ACUTE-PHASE REACTANT THAT SHARES SEQUENCE HOMOLGY WITH OTHER VERTEBRATE LECTINS. *J. Exp. Med.* **167**, 1034–1046 (1988).
50. Turner, M. W. Mannose-binding lectin: The pluripotent molecule of the innate immune system. *Immunol. Today* **17**, 532–540 (1996).
51. Matsushita, M. & Fujita, T. Ficolins and the lectin complement pathway. *Immunol. Rev.* **180**, 78–85 (2001).

52. Fujimori, Y. *et al.* Molecular cloning and characterization of mouse ficolin-A. *Biochem. Biophys. Res. Commun.* **244**, 796–800 (1998).
53. Gorski, J. P. *et al.* Characterization of human C4a anaphylatoxin. *J. Biol. Chem.* **256**, 2707–2711 (1981).
54. Dahl, M. R. *et al.* MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway. *Immunity* **15**, 127–135 (2001).
55. Kishore, U. & Reid, K. B. M. C1q: structure, function, and receptors. *Immunopharmacology* **49**, 159–170 (2000).
56. Strainic, M. G. *et al.* Locally produced complement fragments C5a and C3a provide both costimulatory and survival signals to naive CD4(+) T cells. *Immunity* **28**, 425–435 (2008).
57. Gorski, J. P. *et al.* C4a: the third anaphylatoxin of the human complement system. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5299–5302 (1979).
58. Klos, A. *et al.* The role of the anaphylatoxins in health and disease. *Mol. Immunol.* **46**, 2753–2766 (2009).
59. Guo, R. F. *et al.* Role of C5a-C5aR interaction in sepsis. *Shock* **21**, 1–7 (2004).
60. Huber-Lang, M. *et al.* Generation of C5a by phagocytic cells. *Am. J. Pathol.* **161**, 1849–1859 (2002).
61. Okinaga, S. *et al.* C5L2, a nonsignaling C5A binding protein. *Biochemistry* **42**, 9406–9415 (2003).
62. Flierl, M. A. *et al.* Functions of the complement components C3 and C5 during sepsis. *Faseb J.* **22**, 3483–3490 (2008).
63. Sarma, V. J. *et al.* Complement in lung disease. *Autoimmunity* **39**, 387–394 (2006).
64. Janssen, B. J. C. *et al.* Structure of C3b reveals conformational changes that underlie complement activity. *Nature* **444**, 213–216 (2006).
65. Starkey, P. M. & Barrett, A. J. Evolution of alpha 2-macroglobulin. The demonstration in a variety of vertebrate species of a protein resembling human alpha 2-macroglobulin. *Biochem. J.* **205**, 91–95 (1982).
66. Budd, A. *et al.* Bacterial alpha2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5**, R38 (2004).
67. Doan, N. & Gettins, P. G. W. alpha-Macroglobulins are present in some gram-negative bacteria: characterization of the alpha2-macroglobulin from *Escherichia coli*. *J. Biol. Chem.* **283**, 28747–56 (2008).
68. Robert-Genthon, M. *et al.* Unique features of a *Pseudomonas aeruginosa* α 2-macroglobulin homolog. *MBio* **4**, (2013).
69. Roux, A. *et al.* Combined Inactivation and Expression Strategy To Study Gene Function under Physiological Conditions : Application to Identification of New *Escherichia coli* Adhesins. *J. Bacteriol.* **187**, 1001–1013 (2005).
70. Blaschkowski, H. P. *et al.* Routes of Flavodoxin and Ferredoxin Reduction in

- Escherichia coli*. *Eur. J. Biochem.* **123**, 563–569 (1982).
71. Blattner, F. R. The Complete Genome Sequence of *Escherichia coli* K-12. *Science (80-.)*. **277**, 1453–1462 (1997).
 72. Winsor, G. L. *et al.* Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Res.* **39**, D596–D600 (2011).
 73. Goodman, A. L. *et al.* A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**, 745–754 (2004).
 74. Ventre, I. *et al.* Multiple sensors control reciprocal expression of *Pseudomonas aeruginosa* regulatory RNA and virulence genes. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 171–176 (2006).
 75. Basler, M. & Mekalanos, J. J. Type 6 Secretion Dynamics Within and Between Bacterial Cells. *Science (80-.)*. **337**, 815 (2012).
 76. Schwarz, S. *et al.* What is type VI secretion doing in all those bugs? *TRENDS Microbiol.* **18**, 531–537 (2010).
 77. Neves, D. *et al.* Conformational states of a bacterial α 2-macroglobulin resemble those of human complement C3. *PLoS One* **7**, e35384 (2012).
 78. Pointon, J. a. *et al.* A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. *J. Biol. Chem.* **285**, 33858–33866 (2010).
 79. Mandlik, A. *et al.* Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**, 33–40 (2008).
 80. Telford, J. L. *et al.* Pili in Gram-positive pathogens. *Nat Rev Micro* **4**, 509–519 (2006).
 81. Waksman, G. & Hultgren, S. J. Structural biology of the chaperone–usher pathway of pilus biogenesis. *Nat. rev. Microbiol.* **7**, 765–774 (2009).
 82. Linke-Winnebeck, C. *et al.* Structural model for covalent adhesion of the *Streptococcus pyogenes* pilus through a thioester bond. *J. Biol. Chem.* **289**, 177–189 (2014).
 83. Dodds, A. W. *et al.* The reaction mechanism of the internal thioester in the human complement component C4. *Nature* **379**, 177–179 (1996).
 84. Sauvage, E. *et al.* The penicillin-binding proteins: Structure and role in peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **32**, 234–258 (2008).
 85. Guinane, C. M. *et al.* Contribution of penicillin-binding protein homologs to antibiotic resistance, cell morphology, and virulence of *Listeria monocytogenes* EGDe. *Antimicrob. Agents Chemother.* **50**, 2824–2828 (2006).
 86. Massova, I. & Mobashery, S. MINIREVIEW Kinship and Diversification of Bacterial Penicillin-Binding Proteins and β -Lactamases. *Antimicrob. Agents Chemother.* **42**, 1–17 (1998).
 87. Macheboeuf, P. *et al.* Penicillin binding proteins: key players in bacterial cell cycle

- and drug resistance processes. *FEMS Microbiol. Rev.* **30**, 673–691 (2006).
88. Vollmer, W. *et al.* Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* **32**, 259–286 (2008).
 89. Yousif, S. Y. *et al.* G. Lysis of *Escherichia coli* by β -lactam antibiotics: deletion analysis of the role of penicillin-binding proteins 1A and 1B. *J. Gen. Microbiol.* **131**, 2839–2845 (1985).
 90. Von Rechenberg, M. *et al.* Ampicillin/penicillin-binding protein interactions as a model drug-target system to optimize affinity pull-down and mass spectrometric strategies for target and pathway identification. *Proteomics* **5**, 1764–73 (2005).
 91. Vollmer, W. *et al.* Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of *Escherichia coli*. *J. Biol. Chem.* **274**, 6726–6734 (1999).
 92. Barbosa, M. *et al.* Development of a whole-cell assay for peptidoglycan biosynthesis inhibitors. *Antimicrob. Agents Chemother.* **46**, 943–946 (2002).
 93. Bearne, S. L. & Blouin, C. Inhibition of *Escherichia coli* glucosamine-6 phosphate synthase by reactive intermediate analogues - The role of the 2-amino function in catalysis. *J. Biol. Chem.* **275**, 135–140 (2000).
 94. Badet-Denisot, M. A. *et al.* Characterization of L-glutamine:D-fructose-6-phosphate amidotransferase from an extreme thermophile *Thermus thermophilus* HB8. *Arch. Biochem. Biophys.* **337**, 129–136 (1997).
 95. Bearne, S. L. Active site-directed inactivation of *Escherichia coli* glucosamine-6-phosphate synthase - Determination of the fructose 6-phosphate binding constant using a carbohydrate-based inactivator. *J. Biol. Chem.* **271**, 3052–3057 (1996).
 96. Badet-denisot, M. A. *et al.* MECHANISTIC INVESTIGATIONS ON GLUCOSAMINE-6-PHOSPHATE SYNTHASE. *Bull. Soc. Chim. Fr.* **130**, 249–255 (1993).
 97. Branstrom, A. A. *et al.* In situ assay for identifying inhibitors of bacterial transglycosylase. *Fems Microbiol. Lett.* **191**, 187–190 (2000).
 98. Atrih, A. *et al.* Structural analysis of *Bacillus subtilis* 168 endospore peptidoglycan and its role during differentiation. *J. Bacteriol.* **178**, 6173–6183 (1996).
 99. Benson, T. E. *et al.* The structure of the substrate-free form of MurB, an essential enzyme for the synthesis of bacterial cell walls. *Structure* **4**, 47–54 (1996).
 100. Mengin-Lecreulx, D. *et al.* Identification of the mpl gene encoding UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase in *Escherichia coli* and its role in recycling of cell wall peptidoglycan. *J. Bacteriol.* **178**, 5347–5352 (1996).
 101. Azzolina, B. A. *et al.* The cell wall and cell division gene cluster in the Mra operon of *Pseudomonas aeruginosa*: Cloning, production, and purification of active enzymes. *Protein Expr. Purif.* **21**, 393–400 (2001).
 102. Anderson, M. S. *et al.* Kinetic mechanism of the *Escherichia coli* UDPMurNAc-tripeptide D-alanyl-D-alanine-adding enzyme: Use of a glutathione S-transferase fusion. *Biochemistry* **35**, 16264–16269 (1996).

103. VanHeijenoort, Y. *et al.* MEMBRANE INTERMEDIATES IN THE PEPTIDOGLYCAN METABOLISM OF ESCHERICHIA-COLI - POSSIBLE ROLES OF PBP-1B AND PBP-3. *J. Bacteriol.* **174**, 3549–3557 (1992).
104. Ha, S. H. A. *et al.* The 1.9 Å crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.* **9**, 1045–1052 (2000).
105. Mohammadi, T. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli. *Mol. Microbiol.* **65**, 1106–1121 (2007).
106. Lok-To, S. *et al.* MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science (80-.)*. **719**, 220–222 (2014).
107. Typas, A. *et al.* From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* **10**, 123–36 (2012).
108. Vollmer, W. & Bertsche, U. Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. *Biochim. Biophys. Acta* **1778**, 1714–1734 (2008).
109. Izaki, K., Matsuhara, M. & Stroming, J. L. BIOSYNTHESIS OF PEPTIDOGLYCAN OF BACTERIAL CELL WALLS. 13. PEPTIDOGLYCAN TRANSPEPTIDASE AND D-ALANINE CARBOXYPEPTIDASE - PENICILLIN-SENSITIVE ENZYMATIC REACTION IN STRAINS OF ESCHERICHIA COLI. *J. Biol. Chem.* **243**, 3180–& (1968).
110. Typas, A. *et al.* Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell* **143**, 1097–109 (2010).
111. Hoeltje, J. V. *et al.* U. Novel type of murein transglycosylase in Escherichia coli. *J. Bacteriol.* **124**, 1067–1076 (1975).
112. Contreras-Martel, C. *et al.* Crystal structure of penicillin-binding protein 1a (PBP1a) reveals a mutational hotspot implicated in beta-lactam resistance in Streptococcus pneumoniae. *J. Mol. Biol.* **355**, 684–96 (2006).
113. Bertsche, U. *et al.* Interaction between two murein (peptidoglycan) synthases, PBP3 and PBP1B, in Escherichia coli. *Mol. Microbiol.* **61**, 675–690 (2006).
114. Derouaux, A. *et al.* The monofunctional glycosyltransferase of Escherichia coli localizes to the cell division site and interacts with penicillin-binding protein 3, FtsW, and FtsN. *J. Bacteriol.* **190**, 1831–4 (2008).
115. Hoskins, J. *et al.* Gene disruption studies of penicillin-binding proteins 1a, 1b, and 2a in Streptococcus pneumoniae. *J. Bacteriol.* **181**, 6552–6555 (1999).
116. Vats, P. *et al.* Assembly of the MreB-associated cytoskeletal ring of Escherichia coli. *Mol. Microbiol.* **72**, 170–182 (2009).
117. Wientjes, F. B. & Nanninga, N. ON THE ROLE OF THE HIGH-MOLECULAR-WEIGHT PENICILLIN-BINDING PROTEINS IN THE CELL-CYCLE OF ESCHERICHIA-COLI. *Res. Microbiol.* **142**, 333–344 (1991).
118. Holtje, J. V. Molecular interplay of murein synthases and murein hydrolases in Escherichia coli. *Microb. Drug Resist.* **2**, 99–103 (1996).
119. Lefevre, F. *et al.* Topographical and functional investigation of Escherichia coli

- penicillin-binding protein 1b by alanine stretch scanning mutagenesis. *J. Bacteriol.* **179**, 4761–4767 (1997).
120. Lovering, A. L. *et al.* Structural analysis of an ‘open’ form of PBP1B from *Streptococcus pneumoniae*. *Protein Sci.* **15**, 1701–1709 (2006).
 121. Meisel, U. *et al.* Overproduction of inactive variants of the murein synthase PBP1B causes lysis in *Escherichia coli*. *J. Bacteriol.* **185**, 5342–5348 (2003).
 122. Paik, J. *et al.* Mutational analysis of the *Streptococcus pneumoniae* bimodular class a penicillin-binding proteins. *J. Bacteriol.* **181**, 3852–3856 (1999).
 123. Ho, J. *et al.* Cloning and Characterization of PBP 1C , a Third Member of the Multimodular Class A Penicillin-binding Proteins of *Escherichia coli* *. *J. Biol. Chem.* **274**, 32031–32039 (1999).
 124. Goffin, C. & Ghuysen, J. M. Multimodular penicillin binding proteins: An enigmatic family of orthologs and paralogs. *Microbiol. Mol. Biol. Rev.* **62**, 1079–+ (1998).
 125. Morlot, C. *et al.* Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from *Streptococcus pneumoniae*. *J. Biol. Chem.* **280**, 15984–15991 (2005).
 126. Wissel, M. C. & Weiss, D. S. Genetic analysis of the cell division protein FtsI (PBP3): amino acid substitutions that impair septal localization of FtsI and recruitment of FtsN. *J. Bacteriol.* **186**, 490–502 (2004).
 127. Denome, S. A. *et al.* *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: Viability, characteristics, and implications for peptidoglycan synthesis. *J. Bacteriol.* **181**, 3981–3993 (1999).
 128. Van Asselt, E. J. *et al.* High resolution crystal structures of the *Escherichia coli* lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment. *J. Mol. Biol.* **291**, 877–898 (1999).
 129. Thunnissen, A. *et al.* DOUGHNUT-SHAPED STRUCTURE OF A BACTERIAL MURAMIDASE REVEALED BY X-RAY CRYSTALLOGRAPHY. *Nature* **367**, 750–754 (1994).
 130. Uehara, T. & Park, J. T. Identification of MpaA, and amidase in *Escherichia coli* that hydrolyzes the gamma-D-glutamyl-meso-diaminopimelate bond in murein peptides. *J. Bacteriol.* **185**, 679–682 (2003).
 131. Potluri, L. Septal and lateral wall localization of PBP5, the major D,D-carboxypeptidase of *Escherichia coli*, requires substrate recognition and membrane attachment. *Mol. Microbiol.* **77**, 300–323 (2010).
 132. Fukushima, T. *et al.* Characterization of new L,D-endopeptidase gene product CwlK (previous YcdD) that hydrolyzes peptidoglycan in *Bacillus subtilis*. *Mol. Genet. Genomics* **278**, 371–383 (2007).
 133. Thunnissen, A. *et al.* THE CATALYTIC DOMAIN OF A BACTERIAL LYtic TRANSGLYCOSYLASE DEFINES A NOVEL CLASS OF LYSOZYMES. *Proteins-Structure Funct. Genet.* **22**, 245–258 (1995).
 134. Reid, C. W. *et al.* Inhibition of membrane-bound lytic transglycosylase B by NAG-thiazoline. *Febs Lett.* **574**, 73–79 (2004).

135. Van Straaten, K. E. *et al.* Crystal structure of MltA from *Escherichia coli* reveals a unique lytic transglycosylase fold. *J. Mol. Biol.* **352**, 1068–1080 (2005).
136. Van Asselt, E. J. *et al.* Crystallographic studies of the interactions of *Escherichia coli* lytic transglycosylase Slt35 with peptidoglycan. *Biochemistry* **39**, 1924–1934 (2000).
137. Bateman, A. & Bycroft, M. The structure of a LysM domain from E-coli membrane-bound lytic murein transglycosylase D (MltD). *J. Mol. Biol.* **299**, 1113–1119 (2000).
138. Von Rechenberg, M. *et al.* Affinity chromatography as a means to study multienzyme complexes involved in murein synthesis. *Microb. Drug Resist.* **2**, 155–157 (1996).
139. Izaki, K., Matsuhara, M. & Stroming, J. L. GLYCOPEPTIDE TRANSPEPTIDASE AND D-ALANINE CARBOXYPEPTIDASE - PENICILLIN-SENSITIVE ENZYMATIC REACTIONS. *Proc. Natl. Acad. Sci. U. S. A.* **55**, 656–& (1966).
140. Keck, W. & Schwarz, U. ESCHERICHIA-COLI MUREIN-DD-ENDOPEPTIDASE INSENSITIVE TO BETA-LACTAM ANTIBIOTICS. *J. Bacteriol.* **139**, 770–774 (1979).
141. Keck, W. *et al.* CLONING AND CHARACTERIZATION OF MEPA, THE STRUCTURAL GENE OF THE PENICILLIN-INSENSITIVE MUREIN ENDOPEPTIDASE FROM ESCHERICHIA-COLI. *Mol. Microbiol.* **4**, 209–219 (1990).
142. Kishida, H. *et al.* Crystal structure of penicillin binding protein 4 (dacB) from *Escherichia coli*, both in the native form and covalently linked to various antibiotics. *Biochemistry* **45**, 783–792 (2006).
143. Romeis, T. & Holtje, J. V. SPECIFIC INTERACTION OF PENICILLIN-BINDING PROTEIN-3 AND PROTEIN-7/8 WITH SOLUBLE LYTIC TRANSGLYCOSYLASE IN ESCHERICHIA-COLI. *J. Biol. Chem.* **269**, 21603–21607 (1994).
144. Romeis, T. & Holtje, J. V. PENICILLIN-BINDING PROTEIN-7/8 OF ESCHERICHIA-COLI IS A DD-ENDOPEPTIDASE. *Eur. J. Biochem.* **224**, 597–604 (1994).
145. Harris, F. *et al.* Investigations into the mechanisms used by the C-terminal anchors of *Escherichia coli* penicillin-binding proteins 4, 5, 6 and 6b for membrane interaction. *Eur. J. Biochem.* **269**, 5821–5829 (2002).
146. Heidrich, C. Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol. Microbiol.* **41**, 167–178 (2001).
147. Uehara, T. & Park, J. T. An anhydro-N-acetylmuramyl-L-alanine amidase with broad specificity tethered to the outer membrane of *Escherichia coli*. *J. Bacteriol.* **189**, 5634–5641 (2007).
148. Weidel, W. & Pelzer, H. Bagshaped macromolecules - a new outlook on bacterial cell walls. *Adv. Enzymol.* **26**, 193–232 (1964).
149. SCHLEIFE, K. H. & KANDLER, O. PEPTIDOGLYCAN TYPES OF BACTERIAL CELL-WALLS AND THEIR TAXONOMIC IMPLICATIONS. *Bacteriol. Rev.* **36**, 407–477 (1972).
150. Markiewicz, Z. *et al.* Murein structure and lack of DD- and LD-carboxypeptidase activities in *Caulobacter crescentus*. *J. Bacteriol.* **156**, 649–655 (1983).
151. Glauner, B. *et al.* The composition of the murein of *Escherichia coli*. *J. Biol. Chem.*

- 263**, 10088–10095 (1988).
152. Meroueh, S. O. *et al.* Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4404–4409 (2006).
 153. Gan, L. *et al.* Molecular organization of Gram-negative peptidoglycan. *Proc. Natl Acad. Sci. USA* **105**, 18953–18957 (2008).
 154. Applied, G. *et al.* Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **433**, 531–537 (2005).
 155. Atrih, A. *et al.* Analysis of peptidoglycan structure from vegetative cells of *Bacillus subtilis* 168 and role of PBP 5 in peptidoglycan maturation. *J. Bacteriol.* **181**, 3956–3966 (1999).
 156. Turner, R. D. *et al.* Cell wall elongation mode in Gram-negative bacteria is determined by peptidoglycan architecture. *Nat. Commun.* **4**, 1496 (2013).
 157. Eurofins. Eurofins MWG Operon. (2015). at <<http://www.operon.com/>>
 158. Source. Source BioScience. (2015). at <<http://www.sourcebioscience.com/>>
 159. ExPASy. ProtParam. (2015). at <<http://web.expasy.org/protparam/>>
 160. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 (2005). doi:10.1385/1-59259-890-0:571
 161. Fingerprints facility, U. of D. Proteomic analysis. (2015). at <<http://proteomics.lifesci.dundee.ac.uk/>>
 162. Provencher, S. W. & Glöckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37 (1981).
 163. Whitmore, L. & Wallace, B. a. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
 164. Tucker, H., Allan, W., Greg, D. & Bill, B. SEDNTERP. (2011). at <<http://sednterp.unh.edu/>>
 165. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
 166. Brookes, E. *et al.* The implementation of SOMO (SOLution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: Enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur. Biophys. J.* **39**, 423–435 (2010).
 167. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
 168. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
 169. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
 170. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

171. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 479–485 (2010).
172. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 582–601 (2009).
173. Emsley, P. *et al.* Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
174. Vagin, A. a. *et al.* REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2184–2195 (2004).
175. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
176. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
177. Volkov, V. V & Svergun, D. I. small-angle scattering. 860–864 (2003).
178. Svergun, D. I. *et al.* Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
179. Franke, D. & Svergun, D. I. DAMMIF , a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
180. Svergun, D. *et al.* CRY SOL a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
181. Wong, S. G. & Dessen, A. Supplementary Materials Structure of a bacterial α_2 -macroglobulin reveals mimicry of eukaryotic innate immunity. *Nat. Commun.* 1–10 (2014).
182. Putnam, C. D. *et al.* X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).
183. Kelley, L. A. *et al.* The Phyre2 web portal for protein modeling , prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
184. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science (80-.).* **29**, 997–1003 (2012).
185. Diederichs, K. & Karplus, P. a. Better models by discarding data? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1215–1222 (2013).
186. Vered, M. *et al.* Digestion of elastin. *Int. J. Pept. Protein Res.* 76–84 (1985).
187. Garcia-Ferrer, I. *et al.* Structural and functional insights into *Escherichia coli* α_2 -macroglobulin endopeptidase snap-trap inhibition. *Proc. Natl. Acad. Sci.* 201506538 (2015). doi:10.1073/pnas.1506538112
188. Meyer, C. *et al.* Human α_2 -macroglobulin--another variation on the venus flytrap.

- Angew. Chem. Int. Ed. Engl.* **51**, 5045–7 (2012).
189. Lin, K. *et al.* A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* **21**, 152–159 (2005).
190. Drozdetskiy, A. *et al.* JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* 1–6 (2015). doi:10.1093/nar/gkv332

Publications



Structure of protease-cleaved *Escherichia coli* α -2-macroglobulin reveals a putative mechanism of conformational activation for protease entrapment

Cameron D. Fyfe,^a Rhys Grinter,^a Inokentij Josts,^a Khedidja Mosbahi,^a Aleksander W. Roszak,^{b,c} Richard J. Cogdell,^c Daniel M. Wall,^a Richard J. S. Burchmore,^a Olwyn Byron^d and Daniel Walker^{a*}

Received 24 February 2015

Accepted 30 April 2015

Edited by R. McKenna, University of Florida, USA

Keywords: α -2-macroglobulin; protease inhibitor; conformational change; intrinsic disorder.

PDB reference: *Escherichia coli* α -2-macroglobulin activated by porcine elastase, 4rtf

Supporting information: this article has supporting information at journals.iucr.org/d

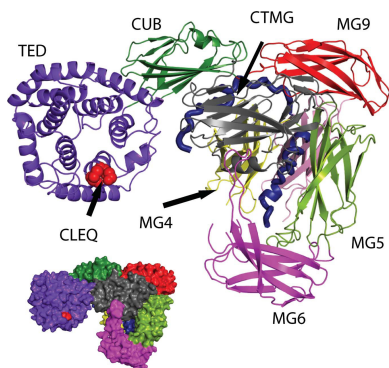
^aInstitute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, ^bWestCHEM, School of Chemistry, College of Science and Engineering, University of Glasgow, Glasgow G12 8QQ, Scotland, ^cInstitute of Molecular Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, and ^dSchool of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland. *Correspondence e-mail: daniel.walker@glasgow.ac.uk

Bacterial α -2-macroglobulins have been suggested to function in defence as broad-spectrum inhibitors of host proteases that breach the outer membrane. Here, the X-ray structure of protease-cleaved *Escherichia coli* α -2-macroglobulin is described, which reveals a putative mechanism of activation and conformational change essential for protease inhibition. In this competitive mechanism, protease cleavage of the bait-region domain results in the untethering of an intrinsically disordered region of this domain which disrupts native interdomain interactions that maintain *E. coli* α -2-macroglobulin in the inactivated form. The resulting global conformational change results in entrapment of the protease and activation of the thioester bond that covalently links to the attacking protease. Owing to the similarity in structure and domain architecture of *Escherichia coli* α -2-macroglobulin and human α -2-macroglobulin, this protease-activation mechanism is likely to operate across the diverse members of this group.

1. Introduction

α -2-Macroglobulins (α 2Ms) are found in eukaryotic blood, invertebrate haemolymph, the eggs of birds and reptiles, and the bacterial periplasm, where they are thought to play a role in the restriction of proteolytic cleavage (Sottrup-Jensen, 1989; Lin *et al.*, 2002; Budd *et al.*, 2004; Li *et al.*, 2004; Doan & Gettins, 2008). Eukaryotic α 2Ms have been shown to play important roles in regulating the proteolytic cleavage of a wide range of proteases and are involved in processes such as fibrinolysis and coagulation (De Boer *et al.*, 1993). Bacterial α 2Ms (BA2Ms) are produced by a wide range of Gram-negative bacteria ranging from human pathogenic and commensal strains to plant pathogens and marine bacteria (Budd *et al.*, 2004). *Escherichia coli* α 2M (ECAM) contains the conserved thioester bond that is characteristic of the α 2Ms and which is essential for covalent binding to cleaving proteases (Budd *et al.*, 2004; Doan & Gettins, 2008; Neves *et al.*, 2012).

The gene encoding ECAM, *yfhM*, is frequently found in an operon with *pbp1C*, which encodes penicillin-binding protein 1C (Pbp1C; Budd *et al.*, 2004; Doan & Gettins, 2008). Pbp1C is predicted to be a bifunctional transpeptidase and transglycosylase owing to its homology to Pbp1A and Pbp1B, which are both essential for the synthesis of the peptidoglycan layer



OPEN ACCESS

(Schiffer & Höltje, 1999; Budd *et al.*, 2004). Both ECAM and Pbp1C are periplasmic proteins that are anchored to the inner membrane and have been proposed to function together in defence and repair against proteases that damage the bacterial cell wall (Budd *et al.*, 2004). Specifically, it has been postulated that the host proteases produced in defence against bacterial infection, which gain access to the periplasmic space, are inhibited by ECAM, with Pbp1C acting to repair damage (Budd *et al.*, 2004).

The overall structure of $\alpha 2$ M comprises a series of β -sheet sandwich macroglobulin (MG) domains forming a 'keyring' shape, a bait-region domain (BRD) spanning the body of the ring, a mostly helical thioester domain (TED) connected to the 'keyring' by a complement protein subcomponent (CUB) domain, and a C-terminal MG (CTMG) domain (Janssen *et al.*, 2005; Marrero *et al.*, 2012; Wong & Dessen, 2014). Within eukaryotic $\alpha 2$ Ms there are eight MG domains including CTMG, which is also known as the receptor-binding domain (MG1–MG7, CTMG), while BA2Ms contain ten MG domains including two N-terminal MG domains anchoring the protein to the inner membrane within the periplasm (MG1–MG9, CTMG) (Huang *et al.*, 1998; Doan & Gettins, 2007; Marrero *et al.*, 2012; Wong & Dessen, 2014). The TED contains a conserved CXEQ motif where a thioester bond is formed, which on activation can covalently link the $\alpha 2$ M to lysine residues on the surface of the attacking protease (Sottrup-Jensen *et al.*, 1980, 1989; Osterberg & Malmsten, 1984; Jacobsen & Sottrup-Jensen, 1993; Janssen *et al.*, 2005; Abdul Ajees *et al.*, 2006; Marrero *et al.*, 2012; Wong & Dessen, 2014). $\alpha 2$ Ms are activated through the protease cleavage of a largely disordered bait region, which results in a conformational change that both traps the protease in a cage-like structure and exposes the highly reactive thioester bond (Travis & Salvesen, 1983; Sottrup-Jensen, 1989; Sottrup-Jensen *et al.*, 1989; Doan & Gettins, 2008).

In human $\alpha 2$ M, chemical cleavage of the thioester bond by methylamine results in a large conformational change that has been interpreted by electron microscopy as similar to that observed on protease cleavage (Sottrup-Jensen *et al.*, 1980; Tapon-Bretaudière *et al.*, 1985; Dodds *et al.*, 1996; Dodds & Law, 1998; Marrero *et al.*, 2012). In contrast, the recently elucidated structures of the BA2M from *Salmonella enterica* serovar Typhimurium (SaA2M) in its unactivated and methylamine-activated forms show that although the overall domain structure of BA2Ms is highly similar to that of human $\alpha 2$ M, there is no major conformational change of the bacterial form on chemical cleavage of the thioester bond (Doan & Gettins, 2008; Neves *et al.*, 2012; Wong & Dessen, 2014). In the structures of both bacterial and eukaryotic $\alpha 2$ Ms the thioester bond lies close to the surface of the TED, but is protected from hydrolysis by a hydrophobic pocket at the interface between the TED and the CTMG domain (Janssen *et al.*, 2005; Le *et al.*, 2012; Wong & Dessen, 2014).

Owing to a lack of detailed structural information on protease-cleaved forms of $\alpha 2$ M, the mechanism through which cleavage of the BRD activates $\alpha 2$ M is not known. However, it has been suggested that upon cleavage within human $\alpha 2$ M the

cleaved BRD interacts with MG2 (MG4 in BA2M), resulting in conformational activation (Marrero *et al.*, 2012). Alternatively, within human $\alpha 2$ M it has been suggested that the bait region interacts with the TED, the MG6 (MG8 in BA2M) and the CUB domains (Marrero *et al.*, 2012). However, in the absence of a crystal structure of a protease-cleaved form of $\alpha 2$ M the mechanism of protease-induced activation remains speculative.

To elucidate the mechanism of protease-induced $\alpha 2$ M activation, we crystallized and solved the X-ray structure of a porcine elastase-cleaved form of ECAM, a close homologue of SaA2M (81% amino-acid sequence identity) for which the structure of the unactivated form was recently solved (Wong & Dessen, 2014). Interestingly, the structure of protease-activated ECAM is highly similar to that of chemically activated human $\alpha 2$ M (12% amino-acid sequence identity) and reveals a clear mechanism of how conformational rearrangement is triggered on protease cleavage. Key to activation is the untethering of the intrinsically disordered bait region on cleavage, allowing this disordered region of polypeptide to outcompete the domain–domain interactions that normally maintain the thioester bond in its unactivated form. This suggests a general mechanism through which members of the large and important $\alpha 2$ M superfamily are activated.

2. Materials and methods

All chemicals were purchased from Sigma unless mentioned otherwise.

2.1. Cloning and protein purification of ECAM

The gene for ECAM, *yfhM* from *E. coli* K-12, was amplified by PCR and cloned into pET-21a vector using NdeI and XhoI restriction sites. The first 22 residues from the N-terminus of the gene, containing a signal sequence identified using *SignalP*, were excluded from the construct. The stop codon was also excluded, resulting in a protein consisting of residues 23–1631 and a C-terminal 6 \times His tag (LEHHHHHH). ECAM was initially overexpressed in *E. coli* BL21(DE3) cells and subsequently in T7 Express Crystal Competent *E. coli* cells (methionine-auxotrophic strain, New England Biosciences) using an inducible T7 promoter with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) as the inducer. Bacteria expressing ECAM were grown in lysogeny broth; selenomethionine-labelled ECAM was obtained using M9 minimal medium supplemented with 50 mg l⁻¹ selenomethionine and 20 mg l⁻¹ of each of nine essential amino acids (excluding methionine). Cells were grown at 37°C to an OD₆₀₀ of 0.6, protein production was induced by the addition of 1 mM IPTG and the cells were grown for a further 6 h. The cell pellet was collected by centrifugation at 4400g for 15 min and the cells were resuspended in binding buffer [20 mM Tris, 10 mM imidazole, 500 mM sodium chloride, 5 mM tris(2-carboxyethyl)phosphine (TCEP) pH 7.5] and lysed by sonication with 1 mg ml⁻¹ lysozyme in the presence of protease inhibitors (Complete Mini, Roche). Cell debris was removed by centri-

Table 1

Data-collection and refinement statistics (single-wavelength anomalous diffraction) for protease-cleaved ECAM.

Data were collected from one crystal. Values in parentheses are for the highest resolution shell.

Data collection	
Space group	<i>H</i> 3
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 176.06, <i>c</i> = 161.13, <i>α</i> = <i>β</i> = 90, <i>γ</i> = 120
Resolution (Å)	46.87–3.65 (4.00–3.65)
Solvent content (%)	69
No. of reflections	20753 (4991)
CC _{1/2}	0.998 (0.675)
<i>R</i> _{merge} (%)	39.0 (378.0)
<i>R</i> _{p.i.m.} † (%)	7.7 (74.4)
<i>I</i> / <i>σ</i> (<i>I</i>)	13.2 (2.1)
Completeness (%)	99.9 (99.9)
Multiplicity	26.6 (26.7)
Anomalous completeness (%)	99.9 (99.3)
Anomalous multiplicity	12.8 (12.8)
DelAnom correlation between half sets	0.335 (0.011)
Mid-slope of anomalous normal probability	1.245
Refinement	
<i>R</i> _{work} / <i>R</i> _{free} (%)	17.7/23.8
No. of atoms	8699
Average <i>B</i> factor (Å ²)	144
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.53
Ramachandran plot‡ (%)	
Favoured	90.7
Allowed	7.9
Outliers	1.3
PDB code	4rtf

† *R*_{p.i.m.} = $\sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

‡ Percentages of residues in favoured/allowed regions calculated by *MolProbity* (Chen *et al.*, 2010).

fugation at 46 000g for 30 min at 4°C. The cell supernatant was then loaded onto a HisTrap HP column (GE Healthcare) and the bound protein was eluted with elution buffer (20 mM Tris, 500 mM imidazole, 500 mM sodium chloride, 5 mM TCEP pH 7.5) using a linear gradient increasing from 10 to 500 mM. Fractions containing ECAM were pooled and dialysed overnight at 4°C into 50 mM Tris, 200 mM sodium chloride pH 7.5 and run on a Superdex S200 gel-filtration column (GE Healthcare). Central fractions from the peak were combined and concentrated using a 100 kDa molecular-weight cutoff centrifugal concentrator.

2.2. Crystallization and structure building

Purified ECAM was reacted in a 1:1 molar ratio with porcine elastase (MP Biomedicals) in 50 mM Tris, 200 mM NaCl pH 7.5 on ice for 5 min before being loaded onto a Superdex S200 gel-filtration column (GE Healthcare). The two major peaks from gel filtration were concentrated to 16 mg ml⁻¹ separately using 100 kDa molecular-weight cutoff centrifugal concentrators and used in crystallization trials. Several hundred crystallization conditions were tested, including the JCSG-plus, MIDAS and Morpheus screens (Molecular Dimensions), for both concentrated peaks. A Cartesian Honeybee 8+1 (Harvard Bioscience) robot was used with 96-well plates, dispensing 0.5 µl reservoir solution and 0.5 µl protein sample. Subsequent scaled-up crystal growth

was performed using 2.5 µl reservoir solution and 2.5 µl protein sample. The initial crystal was grown in conditions consisting of 0.1 M potassium chloride, 0.1 M HEPES, 25% Sokalan CP 7 pH 7.0, and upon optimization the pH was adjusted to 7.5 for larger crystal growth. Crystals were grown using equal volumes of protease-cleaved ECAM and reservoir solution using sitting-drop vapour diffusion, with crystals appearing after 2 d at 16°C for the second fraction and after two weeks for the first fraction. Cryoprotection was optimized with a 3:2 ratio of xylitol-saturated reservoir solution to reservoir solution. Crystals were briefly soaked and flash-cooled in liquid nitrogen for data collection. The best diffraction resolution obtained was 3.8 Å, and molecular replacement with methylamine-activated α2M (PDB entry 4acq) was unsuccessful, most likely as the sequence identity with the human homologue was low (12%) and owing to the difference in domain orientation between the structural models (Marrero *et al.*, 2012). Further expression was performed using a methionine-auxotrophic strain of *E. coli* BL21 (T7 Express Crystal Competent *E. coli*, New England Bioscience) and the purification and crystallization screens were repeated using selenomethionine-labelled protein. As repeating the previous protocol with selenomethionine-labelled protein was unsuccessful, *in situ* proteolytic cleavage screening was performed using porcine elastase. Successful crystallization was achieved using a 1:100 ratio of porcine elastase to selenomethionine-labelled ECAM. Crystallization was successful in the same condition as used previously, with the crystal having a similar appearance and the same space group as previous unlabelled crystals. These crystals diffracted to 3.65 Å resolution and phases were obtained using single-wavelength anomalous diffraction (SAD).

Data were collected for ECAM crystals on the I02, I03 and I24 beamlines at Diamond Light Source, Didcot, England at 100 K at the Se *K* edge (*λ* = 0.97939 Å) using a PILATUS 6M detector. A high-redundancy SAD data set was processed and scaled using *XDS* and *AIMLESS* from the *CCP4* suite of programs (Evans, 2006; Kabsch, 2010; Winn *et al.*, 2011). Selenium sites were located using *SHELXC/D*, with the best substructure solution consisting of 23 sites (Sheldrick, 2010). These selenium sites were input along with the SAD data set to *AutoSol* within the *PHENIX* package to perform phasing and density modification (Terwilliger *et al.*, 2009). Density modification in *AutoSol* was sufficient to break the phase ambiguity owing to the high solvent content of the crystal (69%). This yielded interpretable, low-resolution maps in which density corresponding to secondary-structure elements and larger amino-acid side chains was visible. Initially, the α-helical TED domain was built in *Coot* using idealized α-helical sections (Emsley *et al.*, 2010). The loops between these sections were connected where density was available. The six selenomethionine sites in this domain provided the starting sites for the building of amino-acid side chains in the TED domain, although owing to the resolution initially only larger side chains and those where continuous sequence could be determined were built. In addition to the building of the TED domain, a number of α2M-derived polyalanine MG

domains were rigid-body fitted into their corresponding density and manually real-space refined in *Coot* (Emsley *et al.*, 2010). As with the TED domain, where possible side chains

were fitted using the positions of selenomethionine Se atoms as starting sites. This initial building yielded a partial model, which was then used in conjunction with the selenomethionine

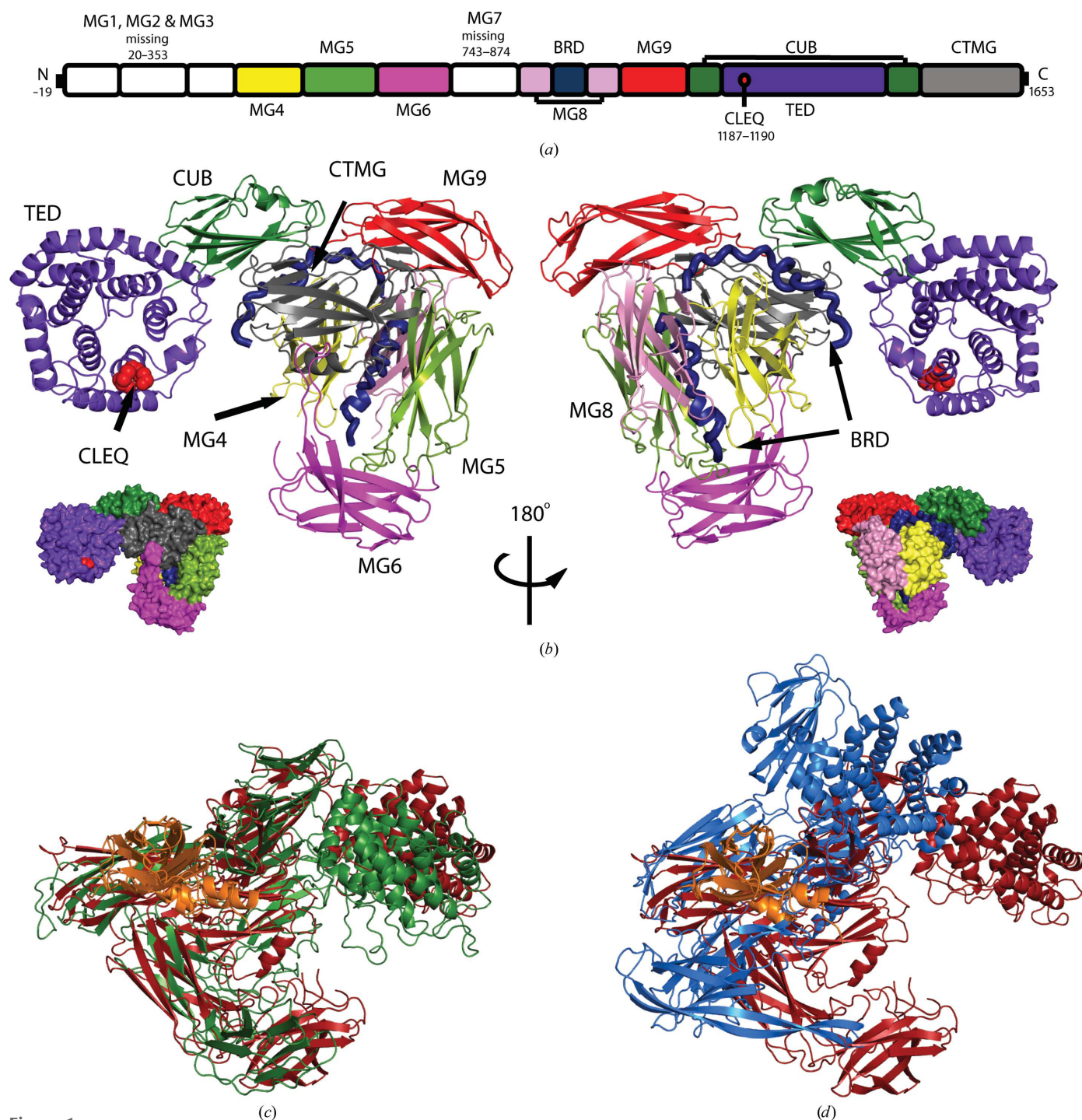


Figure 1

Crystal structure of porcine elastase-cleaved *E. coli* α2M (ECAM). (a) Schematic representation of the 13 domains of ECAM showing the macroglobulin domains (MG) including the C-terminal MG (CTMG) domain, the bait-region domain (BRD), the complement protein subcomponent (CUB) domain and the thioester domain (TED) containing the CLEQ motif. (b) The structure of elastase-cleaved ECAM with the individual domains coloured as in (a). Smaller van der Waals surfaces in both views are also presented. In the left view the α-helical TED is orientated with the CLEQ thioester (drawn as red van der Waals spheres) positioned above the pocket which is thought to accommodate the attacking protease. (c) Structural alignment of methylamine-treated human α2M (PDB entry 4acq monomer trimmed to the domains present in cleaved ECAM) and elastase-cleaved ECAM (PDB entry 4rtd) in green and red, respectively. (d) Structural alignment of native SaA2M (PDB entry 4u48 trimmed to the domains present in cleaved ECAM) and elastase-cleaved ECAM (PDB entry 4rtd) in blue and red, respectively. Structural alignments were performed using the MG domain containing the bait region (shown in orange).

substructure to rephase the experimental data using MR-SAD phasing in *Phaser* (McCoy *et al.*, 2007). This process led to phase improvement and the appearance of new features in the map, which were modelled, and the process was repeated iteratively. Partway through the building process the atomic coordinates for SaA2M were published (PDB entry 4u48; Wong & Dessen, 2014), and the domains from this model provided validation of the MG-domain placement and side-chain modelling in our experimentally phased map. The SaA2M structure also provided a template for building the more difficult sections of the model. At this point restrained TLS refinement using *REFMAC5* was found to stably improve both R_{work} and R_{free} , and refinement was performed and the model was improved and finished manually in *Coot* (Murshudov *et al.*, 2011; Emsley *et al.*, 2010). Electron density for the thioester bond indicated that the deaminated glutamine forms no covalent bond to the cysteine. Before submission of the final model, the quality of the structure was assessed using the *MolProbity* webserver (Chen *et al.*, 2010). The atomic coordinates and structure factors were deposited in the Protein Data Bank (PDB entry 4rtnd). Statistics for data collection, experimental phasing and refinement are presented in Table 1. For mass-spectrometric analysis of protease-cleaved ECAM, crystals were washed in reservoir solution before being dissolved in deionized water and heated to 96°C in bromophenol blue sample buffer for 5 min. The sample was then run on a NuPAGE Novex 4–12% bis-tris gel (Invitrogen) and visible bands were cut for proteomic analysis. Samples were digested by trypsin and analysed by LC-MS/MS (Orbitrap XL) performed at the Fingerprints Proteomics Facility at the University of Dundee.

3. Results

3.1. Overall structure of protease-activated ECAM

To determine the structural changes that occur on protease cleavage of ECAM, we performed protease digestion with porcine elastase and used the major products from gel filtration of cleaved ECAM to perform crystallization trials. This yielded diffracting crystals in 0.1 *M* potassium chloride, 0.1 *M* HEPES, 25% Sokalan CP 7 pH 7.0 with cleaved ECAM. In order to obtain phase information, we attempted to repeat this process with selenomethionine-labelled ECAM, but this failed to yield crystals. However, an alternative strategy of *in situ* proteolysis and crystallization with selenomethionine-labelled ECAM was successful. This method yielded crystals that diffracted to 3.65 Å resolution. Upon completion and validation of the model of protease-activated ECAM, the number of Ramachandran outliers remaining was appropriate for a crystal structure with a resolution of 3.65 Å. Although the R_{merge} and $R_{\text{p.i.m.}}$ values were high, this can be explained by the highly redundant data set used; with a $CC_{1/2}$ of 0.675 in the highest shell, these data used were judged to be acceptable (Karplus & Diederichs, 2012; Diederichs & Karplus, 2013).

Similar to the domain architecture of native SaA2M, uncleaved ECAM consists of ten MG domains, with a largely

disordered bait-region domain found within MG8, a TED that houses the reactive thioester bond and a CUB domain (Fig. 1*a*). Elastase-cleaved ECAM (Fig. 1*b*) adopts a conformation similar to that of methylamine-activated human $\alpha 2\text{M}$ (Fig. 1*c*) but distinct from the unactivated form of SaA2M (Fig. 1*d*). Despite the low sequence identity between human $\alpha 2\text{M}$ and ECAM (12%), the r.m.s.d. between C^α atoms for these proteins is 14.1 Å, while that for SaA2M and ECAM, which share 82% sequence identity, is 22.1 Å. Most notably, in the structure of the protease-cleaved ECAM, interactions between the TED and the CTMG domain, which protects the thioester bond in unactivated SaA2M, are not present. Instead, the thioester region of TED is solvent-exposed and faces the expected location of the attacking protease, as it would be positioned when cleaving the bait region of ECAM (Fig. 1*b*). Electron density for elastase, in addition to that for MG domains 1, 2, 3 and 7, was absent in $2F_o - F_c$ maps of cleaved ECAM, as was electron density for 20 amino-acid residues (Arg923–Leu942) within the bait region. Analysis of elastase-cleaved ECAM by SDS–PAGE and mass spectrometry confirms that MG domains 1, 2, 3 and 7 are absent from

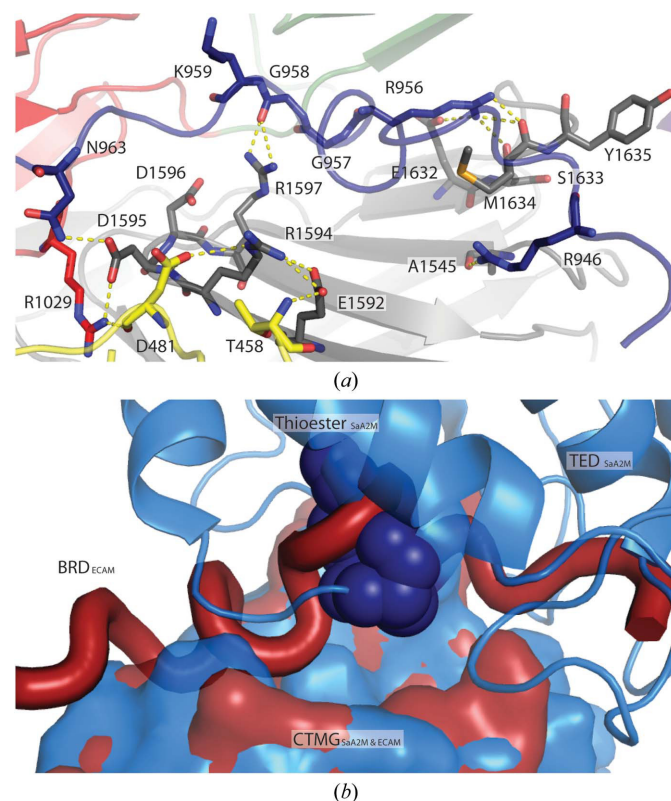


Figure 2

Bait-region interaction with the CTMG domain. (*a*) Here we show hydrogen bonding between the cleaved bait region (blue) and the CTMG domain (grey) as dashed yellow lines. Note the interaction between Arg956 and Met1634, with the CTMG methionine normally involved in the hydrophobic pocket found in uncleaved native SaA2M. MG4 and MG9 are shown in yellow and red, respectively. (*b*) Surface representation of the CTMG domain bound to the TED in SaA2M (blue) overlaid (by superposition of CTMG domains) with the previously disordered bait region bound to the CTMG domain in elastase-cleaved ECAM (red). The thioester bond within the TED is shown as dark blue van der Waals spheres.

the crystallized protein (Supplementary Fig. S1). Two bands at 90 and 75 kDa on SDS–PAGE have similar peptide coverage, encompassing the same domains (Supplementary Fig. S1). However, the reason for the difference in apparent molecular weights between these two species is not known.

3.2. Interaction of the BRD with the CTMG domain

Although it is known that protease cleavage of the largely disordered bait region of both human and bacterial α 2Ms gives rise to a large conformational rearrangement and activation of the thioester bond, the mechanism through which this is mediated was unclear (Barrett *et al.*, 1979; Neves *et al.*, 2012). However, the structure of protease-cleaved ECAM provides a clear mechanism for this process. Thioester bond release is achieved by the untethering of a large region of an unstructured polypeptide chain upon cleavage of the BRD that forms new interactions with the CTMG domain, thereby preventing the protective interaction with the TED. Specifically, in the elastase-cleaved form of ECAM additional residues (Phe947–Asn963) of the BRD are observed that are disordered in the uncleaved SaA2M structure. All of these residues in the protease-cleaved ECAM structure are ordered owing to the formation of a new binding interface between the BRD and the CTMG domain upon protease cleavage (Fig. 2a). Critically, the region of the CTMG domain that is involved in BRD binding substantially overlaps with the region of the CTMG domain that forms the binding interface with the TED in the uncleaved form of the protein (Fig. 2b). The buried surface area between the CTMG domain and TED in SaA2M is 1972 Å², with three hydrogen bonds between these domains, whereas the buried surface area between the elastase-cleaved BRD and the CTMG domain in ECAM is 1438 Å², with five hydrogen bonds between the domains. In the protease-cleaved ECAM, hydrogen-bond formation between residues of the CTMG domain and the cleaved bait region involves the highly conserved RDDR and EXMY motifs (Fig. 3). Interestingly, it is residues within these motifs

that also form hydrogen bonds to the TED in the uncleaved SaA2M protein (Fig. 3; Wong & Dessen, 2014).

3.3. Cleavage-induced conformational changes and thioester-bond activation of α 2Ms

We propose that the loss of the interaction of the TED with the CTMG domain is sufficient to enable both the large TED movement observed in the cleaved ECAM structure relative to the unactivated SaA2M structure and the exposure of the thioester bond (Fig. 4a, Supplementary Movie S1). The movement of the TED domain shows an overall shift of 36 Å, with the MG6 domain moving by 50 Å and with both the MG6 and TED arms hugging the position at which the attacking protease would be located to cleave the BRD (Fig. 4a, Supplementary Movie S1). In addition to these global conformational changes, protease cleavage leads to localized changes in the environment of the thioester bond that are likely to lead to its activation. In the unactivated SaA2M structure the conserved methionine and tyrosine side chains (Met1625 and Tyr1626 in SaA2M) from the CTMG domain form part of the hydrophobic pocket at the interface with the TED that has been shown to be important in maintenance of the thioester bond (Wong & Dessen, 2014). The loss of these key side-chain interactions and the movement of the additional side chains which constitute the protective hydrophobic pocket (Tyr1175 and Tyr1177 in SaA2M and Tyr1183 and Tyr1185 in ECAM) from the TED exposes the thioester bond to the solvent, allowing hydrolysis or covalent-bond formation with the attacking protease (Fig. 4b). As there are no contacts present from the CTMG domain providing a hydrophobic pocket to protect the thioester from hydrolysis by water, the thioester bond could be cleaved and the deaminated glutamine (Gln1190) would be converted to a glutamic acid. Within our protease-activated ECAM structure the distance between the S atom of Cys1187 and the C γ atom of Gln1190 is 4.6 Å, indicating that the thioester bond may not be intact (Supplementary Fig. S2). However, we cannot rule out the possibility that there is a mixed population of molecules, some of which

TED					CTMG					
Escherichia coli A2M (yfhM)	P76578	1181	KAYFYGCLEQTA	1192	1592	EFRDGRFV	1599	1630	MVESMYVP	1637
Salmonella typhimurium	Q8ZN46	1173	KAYFYGCLEQTT	1184	1583	EFRDGRFV	1590	1621	QVESMYAP	1628
Pseudomonas putida	Q88QC4	1160	QAYFYGCLEQTT	1171	1571	EYRDGRYV	1578	1609	QVESMYAP	1616
Bordetella pertussis	Q7VVC2	1222	LTIFYGCTEQTI	1233	1663	EFRDGRYV	1670	1700	EAEDEMYRP	1707
Desulfovibrio alaskensis	Q316S	1188	DRYFYGCLEQTA	1199	1582	LRQDGRVV	1590	1629	SAEDMYDP	1636
Rhizobium meliloti	Q92VA6	1361	DRYFYGCLEQTT	1372	1751	EFRDGRFV	1758	1791	NVEDMYRP	1798
Helicobacter hepaticus	Q7VI74	1371	IRYFYGCLEQTT	1382	1755	DIEEDKIM	1762	1792	NAEDMYDN	1799
Fusobacterium nucleatum	D5RAI8	1159	LDYFYICLEQIS	1170	1558	DIEEDRVA	1565	1596	KVESMYNN	1603
Rickettsia conorii	Q92HD6	1435	DNYFYGCTEQLI	1446	1834	NNEDGRVM	1841	1871	YSEEDMYDP	1878
Leptospira interrogans	M5VM24	1485	NRYFYSCLEQKL	1496	1863	LTKEDENWW	1870	1915	RVEDMYLP	1922

Figure 3

Sequence alignments of conserved motifs involved in the thioester pocket and conformational activation. Highlighted in yellow are tyrosines important for maintaining the hydrophobic pocket protecting the thioester bond (CLEQ motif, also shown in yellow). The tyrosine within the C-terminal macroglobulin (CTMG) domain, involved in protecting the thioester, which coordinates to the glutamate within the CLEQ motif in the native SaA2M structure is highlighted in blue (PDB entry 4u48). The conserved proline found near the thioester is highlighted in green along with the arginine that it coordinates in unactivated SaA2M. The residues highlighted in pink, orange and red in the CTMG domain coordinate to residues Asn963, Gly958 and Arg956, respectively, in the cleaved bait region of elastase-cleaved ECAM.

possess an intact thioester bond. Owing to this ambiguity, we have represented the thioester without a covalent bond between Cys1187 and the C γ atom of Gln1190 and have also omitted the O atom from the deaminated glutamine that would be formed on hydrolysis of the thioester bond.

4. Discussion

Although it has been suggested that small-molecule-activated human α 2M resembles the conformation of the protease-cleaved form, there have to date been no detailed structural data to confirm this (Tapon-Breaudière *et al.*, 1985). The similarity in the overall conformation of protease-cleaved ECAM to that of methylamine-activated human α 2M confirms that the protease-activated and chemically activated forms of human α 2M are structurally equivalent. In addition, these data suggest that the entrapment of cleaving proteases is likely to occur in a similar manner for BA2Ms as has been proposed for human α 2M (Marrero *et al.*, 2012; Meyer *et al.*, 2012). However, there are clearly key differences in the details of the interactions that maintain the inactive conformations of BA2Ms and human α 2M, since the recent structures of unactivated and methylamine-activated SaA2M show that chemical cleavage of the thioester bond does not lead to global conformational changes in this case (Marrero *et al.*, 2012; Wong & Dessen, 2014). This difference is likely to be owing to the domain location of the side chains that comprise the thioester-protecting pocket (Supplementary Fig. S3). In BA2Ms this pocket comprises two tyrosine side chains from the TED and a tyrosine and a methionine side chain from the

CTMG domain, whereas in eukaryotic α 2Ms all four residues (three tyrosine side chains and one methionine side chain) are found in the CTMG domain. When SaA2M is reacted with methylamine no conformational change is seen, but Tyr1175 from the TED is displaced (Wong & Dessen, 2014). The structural counterpart of this side chain (Tyr1307 from TEP1) in eukaryotic α 2M family members is, however, found within the CTMG domain (Supplementary Fig. S3; Baxter *et al.*, 2007). It is presumably the rearrangement of this side chain and perhaps other CTMG side chains that comprise the thioester-protecting pocket which leads to the loss of TED–CTMG domain binding and subsequent global conformational changes. The buried surface area seen between the TED and the CTMG domain in native SaA2M decreases by 27% compared with that of the BRD and the CTMG domain in protease-cleaved ECAM; however, the number of hydrogen bonds increases. Although there is a decrease in the buried surface area, it is the interaction between important residues for protecting the thioester that would be expected to trigger conformational change releasing TED and allow subsequent interaction with a cleaving protease.

The role of BA2Ms, which are inner membrane-anchored periplasmic proteins, has been suggested as protease inhibitors that inhibit exogenous proteases that have breached the outer membrane. Comparison of the structures of unactivated SaA2M and protease-cleaved ECAM illustrates how protease-induced conformational changes enable protease entrapment (Fig. 5). The movement of the TED-domain arm and MG6-domain arm around the central pocket above the bait region is similar to the movement seen in methylamine-activated

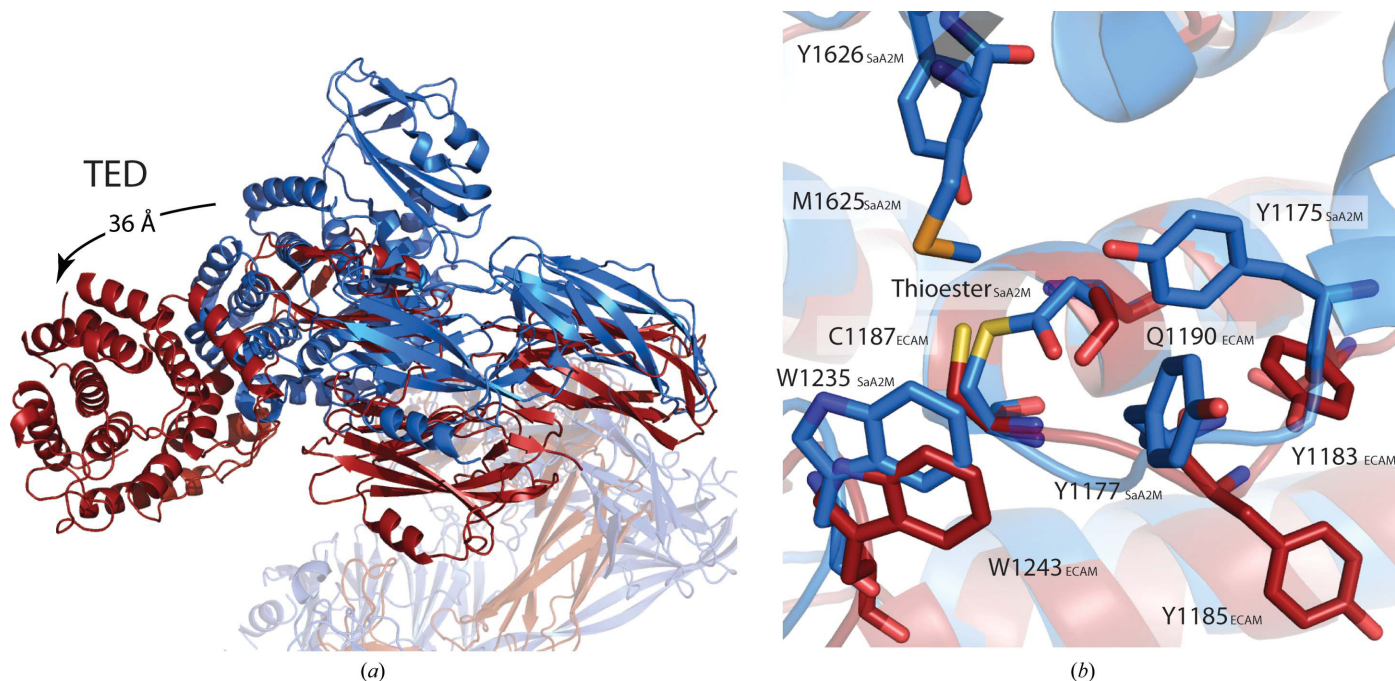


Figure 4

Conformational shift of the TED and thioester of protease-activated ECAM. (a) Superposition of protease-cleaved ECAM (red) and SaA2M (blue) showing a 36 Å shift of the TED. Structures were overlaid by superposition of MG8 containing the BRD. (b) Overlaid elastase-cleaved ECAM (red) and native SaA2M (blue) TEDs. Tyrosines thought to protect the thioester region from hydrolysis in native SaA2M (Tyr1175 and Tyr1177) are orientated away from the protease-activated ECAM thioester region (Tyr1183 and Tyr1185 in protease-activated ECAM).

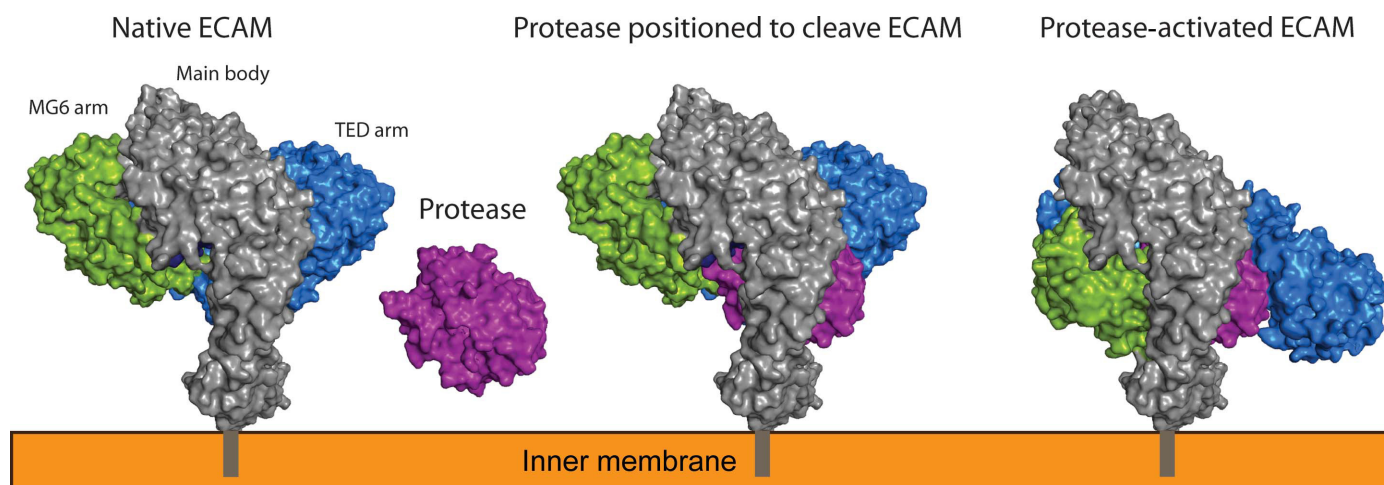


Figure 5

Putative mechanism of protease entrapment and inhibition by ECAM. ECAM with a membrane anchor within the periplasm of bacteria encounters protease and forms a covalent complex upon cleavage, inhibiting the protease from the cleavage of large substrates. The main body of ECAM anchored to the inner membrane is shown in grey and contains the domains showing little movement (MG4, MG8 and BRD) between native and protease-activated structures or that are not present in the protease-activated structure (MG1–MG3 and MG7). The blue TED arm (TED, CUB, CTMG and MG9 domains) and green MG6 arm (MG5 and MG6 domains) entrap the protease with the blue arm containing the thioester that forms a covalent bond.

human $\alpha 2M$; when accounting for the domains that are not present in the protease-activated ECAM structure no clashes are seen between the moving arms and the absent domains (Marrero *et al.*, 2012). The entrapment of proteases would limit the proteolysis to smaller substrates, as has been suggested for human $\alpha 2M$, and would prevent the cleavage of important larger substrates such as the peptide component of the peptidoglycan layer or large proteins (Fig. 5; Sottrup-Jensen, 1989). Although we cannot be sure why the covalently bound elastase is not present in our structure, this may be owing to a lack of available and correctly positioned lysine side chains on the surface of the protease, since the thioester bond is preferentially cleaved by this side chain (Marrero *et al.*, 2012). The lack of MG domains 1, 2, 3 and 7 within the crystal lattice is likely to be owing to the MG domains being cleaved by the elastase.

In summary, the structure of protease-activated ECAM shows a competitive mechanism of activation in which cleavage of the BRD allows the normally intrinsically disordered region of this domain to outcompete the TED for CTMG-domain binding. Loss of the TED–CTMG domain interaction leads to a large conformational rearrangement of ECAM and exposure of the reactive thioester bond. The structural similarity between methylamine-activated human $\alpha 2M$ and protease-cleaved ECAM suggests that similar mechanisms are likely to operate across the diverse members of the $\alpha 2M$ family.

Acknowledgements

We would like to acknowledge the Diamond Light Source for access to beamlines I02, I03 and I24 (proposal No. MX5689). CDF is supported by a four-year studentship from the BBSRC/EP SRC doctoral training centre in cell and proteomic engineering (Grant No. EP/F500424/1). IJ is supported by a four-year studentship from the Wellcome Trust (Grant No.

093597/Z/10/Z). RG is supported by a Kelvin Smith Scholarship from the University of Glasgow.

References

- Abdul Ajees, A., Gunasekaran, K., Volanakis, J. E., Narayana, S. V. L., Kotwal, G. J. & Krishna Murthy, H. M. (2006). *Nature (London)*, **444**, 221–225.
- Barrett, A. J., Brown, M. A. & Sayers, C. A. (1979). *Biochem. J.* **181**, 401–418.
- Baxter, R. H. G., Chang, C.-I., Chelliah, Y., Blandin, S., Levashina, E. A. & Deisenhofer, J. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 11615–11620.
- Boer, J. P. de, Creasey, A. A., Chang, A., Abbink, J. J., Roem, D., Eerenberg, A. J., Hack, C. E. & Taylor, F. B. (1993). *Infect. Immun.* **61**, 5035–5043.
- Budd, A., Blandin, S., Levashina, E. A. & Gibson, T. J. (2004). *Genome Biol.* **5**, R38.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
- Doan, N. & Gettins, P. G. W. (2007). *Biochem. J.* **407**, 23–30.
- Doan, N. & Gettins, P. G. W. (2008). *J. Biol. Chem.* **283**, 28747–28756.
- Dodds, A. W. & Law, S. K. A. (1998). *Immunol. Rev.* **166**, 15–26.
- Dodds, A. W., Ren, X.-D., Willis, A. C. & Law, S. K. A. (1996). *Nature (London)*, **379**, 177–179.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Huang, W., Dolmer, K., Liao, X. & Gettins, P. G. W. (1998). *Protein Sci.* **7**, 2602–2612.
- Jacobsen, L. & Sottrup-Jensen, L. (1993). *Biochemistry*, **32**, 120–126.
- Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B. & Gros, P. (2005). *Nature (London)*, **437**, 505–511.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Le, B. V., Williams, M., Logarajah, S. & Baxter, R. H. G. (2012). *PLoS Pathog.* **8**, e1002958.
- Li, Z.-F., Wu, X.-H. & Engvall, E. (2004). *Genomics*, **83**, 1083–1093.

- Lin, M., Sutherland, D. R., Horsfall, W., Totty, N., Yeo, E., Nayar, R., Wu, X.-F. & Schuh, A. C. (2002). *Blood*, **99**, 1683–1691.
- Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G. R., Navaza, J., Sottrup-Jensen, L. & Gomis-Rüth, F. X. (2012). *Angew. Chem. Int. Ed.* **51**, 3340–3344.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Meyer, C., Hinrichs, W. & Hahn, U. (2012). *Angew. Chem. Int. Ed.* **51**, 5045–5047.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Neves, D., Estrozi, L. F., Job, V., Gabel, F., Schoehn, G. & Dessen, A. (2012). *PLoS One*, **7**, e35384.
- Osterberg, R. & Malmensten, B. (1984). *Eur. J. Biochem.* **143**, 541–544.
- Schiffer, G. & Höltje, J. V. (1999). *J. Biol. Chem.* **274**, 32031–32039.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Sottrup-Jensen, L. (1989). *J. Biol. Chem.* **264**, 11539–11542.
- Sottrup-Jensen, L., Petersen, T. E. & Magnusson, S. (1980). *FEBS Lett.* **121**, 275–279.
- Sottrup-Jensen, L., Sand, O., Kristensen, L. & Fey, G. H. (1989). *J. Biol. Chem.* **264**, 15781–15789.
- Tapon-Bretaudière, J., Bros, A., Couture-Tosi, E. & Delain, E. (1985). *EMBO J.* **4**, 85–89.
- Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H. & Hung, L.-W. (2009). *Acta Cryst.* **D65**, 582–601.
- Travis, J. & Salvesen, G. S. (1983). *Annu. Rev. Biochem.* **52**, 655–709.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Wong, S. G. & Dessen, A. (2014). *Nature Commun.* **5**, 4917.